

UNIVERSITEIT ANTWERPEN
FACULTEIT WETENSCHAPPEN
DEPARTEMENT FYSICA

MASTERTHESIS

Deep learning voor automatische materiaalidentificatie in elektronen energieverlies spectroscopie

Student:
Arno Annys

Promotor:
Prof. Dr. Johan Verbeeck
Co-promotor:
Dr. Daen Jannis

Mei 2023

Proefschrift voorgelegd tot het behalen van de graad van master in de fysica aan de
Universiteit Antwerpen, te verdedigen door Arno Annys



Abstract

Electron energy loss spectroscopy (EELS) is a well established technique in electron microscopy that yields information on the elemental content of a sample in a very direct manner. One of the persisting limitations of EELS is the requirement for manual identification of core-loss edges and their corresponding elements. This can be especially bothersome in spectrum imaging. This thesis introduces a synthetic dataset with 736 000 labeled EELS spectra, computed from available generalized oscillator strength tables, that represents 107 K, L, M or N core-loss edges in 80 chemical elements. Generic lifetime-broadened peaks and oscillations are used to mimic the fine structure observed in experimental core-loss edges. The proposed dataset is used to train and evaluate a series of neural network architectures, being a multilayer perceptron, a convolutional neural network, a U-Net, a residual neural network, a vision transformer and a compact convolutional transformer. The resulting neural networks can successfully identify elements in a broad range of experimental spectra. Finally, an ensemble of neural networks is used to demonstrate fully automated elemental mapping in spectrum images, either by directly mapping the predicted elemental content or by using the predicted content as input for a model based mapping.

Samenvatting

Elektronen energieverlies spectroscopie (EELS) is een populaire techniek in elektronenmicroscopie die op zeer directe wijze informatie levert over de elementaire inhoud van een sample. Eén van de voornaamste beperkingen van EELS is de vereiste tot manuele identificatie van core-loss edges en hun overeenkomstige elementen. Dit kan in het bijzonder problematisch zijn in spectrale beelden. Deze thesis introduceert een synthetische dataset met 736 000 gelabelde EELS spectra, berekend a.d.h.v. beschikbare generalised oscillator strength tabellen. De dataset representeert 107 K, L, M of N core-loss edges en 80 chemische elementen. Generieke levensduur-verbrede pieken en oscillaties worden gebruikt voor het nabootsen van de fijnstructuur die voorkomt voor experimentele edges. De voorgestelde gesimuleerde dataset wordt gebruikt voor het trainen en evalueren van een reeks neurale netwerk architecturen, namelijk een meerlaags perceptron, een convolutioneel neuraal netwerk, een U-Net, een residueel neuraal netwerk, een vision transformer en een compacte convolutionele transformer. De resulterende neurale netwerken slagen er in elementen te identificeren in een breed gamma aan experimentele spectra. Ten slotte wordt een ensemble van neurale netwerken gebruikt om een volledig geautomatiseerde elementaire mapping in spectrale beelden te demonstreren, zowel door rechtstreeks de voorspelde elementaire inhoud te mappen als door de voorspellingen te gebruiken als automatische input voor een model-gebaseerde mapping.

Inhoudsopgave

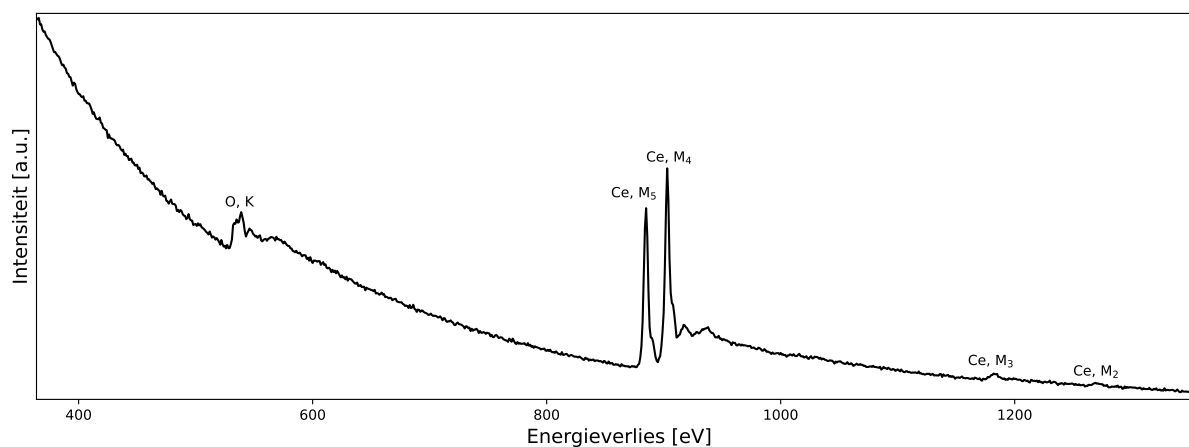
1	Inleiding	1
1.1	Elektronen energieverlies spectroscopie	1
1.2	Artificiële intelligentie en deep learning	2
1.3	Kader van de studie en gerelateerd werk	5
2	Vorming van de dataset	9
2.1	EELS theorie	9
2.1.1	Core-loss excitaties	9
2.1.2	Low-loss gebied	10
2.1.3	Fijnstructuur	11
2.2	Synthetische dataset	12
2.2.1	Inhoud van de dataset	12
2.2.2	Werkwijze van de simulatie	13
2.3	Experimentele dataset	24
2.4	Verdeling van training-, validatie- en testset	24
3	Neurale netwerken voor elementidentificatie	27
3.1	Netwerk architecturen	27
3.2	Training en optimalisatie	33
3.3	Resultaten	36
3.3.1	Dimensiereductie-methode	36
3.3.2	Vergelijking van architecturen	37
3.3.3	Graad van overfitting	39
3.3.4	Ensemble van neurale netwerken	39
3.3.5	Verwarringsmatrix	41
3.3.6	Uitsluiting van N edges	43
3.3.7	Kwalitatieve evaluatie	43
3.3.8	Grad-CAM	45
3.3.9	Randgevallen	46
4	Toepassing: Automatische element-mapping	47
4.1	LaMnO ₃ /BaTiO ₃ /SrTiO ₃ superrooster	48
4.2	Ti-O-N sample	49

5	Bespreking	52
6	Conclusie en verder onderzoek	54
A	Appendix	63
A.1	Benoemingconventies van core-loss edges	63
A.2	Evaluatie van extra U-Net, ViT en CTT instanties.	64
A.3	Voorspellingen op experimentele spectra	65

Inleiding

1.1 Elektronen energieverlies spectroscopie

Elektronen energieverlies spectroscopie (EELS) is een analytische techniek in elektronenmicroscopie gebruikt voor de studie van de elektronische eigenschappen van een materiaal. EELS is gebaseerd op het analyseren van de energieverdeling van een initieel monochromatische elektronenbundel, nadat deze geïnterageerd heeft met een sample. EELS levert op zeer directe wijze informatie over de elementaire inhoud van een sample. Deze informatie zit vervat in de zgn. ionisatie pieken, die het gevolg zijn van de ionisatie van atomaire elektronen door snelle elektronen uit de bundel. Het energieverlies van een snel elektron is karakteristiek voor het atoom dat het ioniseert. EELS spectra worden verdeeld in twee groepen, bepaald door het energieverliesgebied dat wordt weergegeven. Low-loss spectra omvatten een energieverlies van nul tot ongeveer 100 eV. Een low-loss spectrum bestaat o.a. uit de zero-loss piek en plasmon pieken. De zero-loss piek is het gevolg van bundel-elektronen die niet inelastisch verstrooid zijn door het sample. Plasmon pieken, welke het gevolg zijn van collectieve excitaties van de conductie-elektronen, dragen informatie over de diëlektrische eigenschappen, vorm en dikte van het sample. Core-loss spectra omvatten het energieverliesgebied hoger dan het low-loss gebied. De ionisatie pieken in het core-loss gebied worden core-loss edges genoemd. De ionisatie energie van een atomair elektron vormt het energieverlies waarbij de overeenkomstige core-loss edge start, en wordt de edge onset energie genoemd. Core-loss edges worden gelabeld als K, L, M, N of O volgens het hoofdkwantumgetal van het geïoniseerde orbitaal. Deze labels worden verder geïndexeerd i.f.v. het orbitaal en totaal impulsmoment. Een tabel met de benoemingconventies van EELS edges is gegeven in appendix A.1. De splitsing van energieniveaus maakt dat er gescheiden sub-edges voorkomen in het spectrum, welke geclassificeerd worden als major of minor edges. Figuur 1.1 toont een typisch voorbeeld van een core-loss EELS spectrum waarop de zuurstof K edge en de cerium M edges te zien zijn.



Figuur 1.1. Core-loss EELS spectrum van CeO_2 afkomstig van EELS.info [1]

In scanning transmissie elektronenmicroscopie (STEM) levert EELS grote hoeveelheden data in de vorm van spectrale beelden. Elementidentificatie in elk individueel spectrum, m.a.w. bij elke probe-positie, laat een mapping toe van de ruimtelijke verdeling van elementen in het sample. In EELS spectra met voldoende hoge energie-resolutie kan – uit een analyse van de zgn. fijnstructuur van een core-loss edge – informatie over de elektronische toestand verkregen worden. Dit laat toe om bv. oxidatie-toestand of chemische binding te bestuderen. Naast een kwalitatieve studie van EELS spectra, zoals elementidentificatie en mapping, is ook een kwantitatieve studie mogelijk. Een kwantificatie van core-loss edges laat toe relatieve en absolute hoeveelheden van elementen in het sample te bepalen. Origineel gebeurde deze kwantificatie d.m.v. integratie van core-loss edges nadat deze werden geïsoleerd van de spectrum-achtergrond. Recenter gebeurt kwantificatie via een model-gebaseerde aanpak, waarbij een achtergrond en de cross-sections van de aanwezige elementen aan het spectrum gefit worden [2]. Deze model-gebaseerde aanpak vereist als input de aanwezige elementen in het spectrum.

1.2 Artificiële intelligentie en deep learning

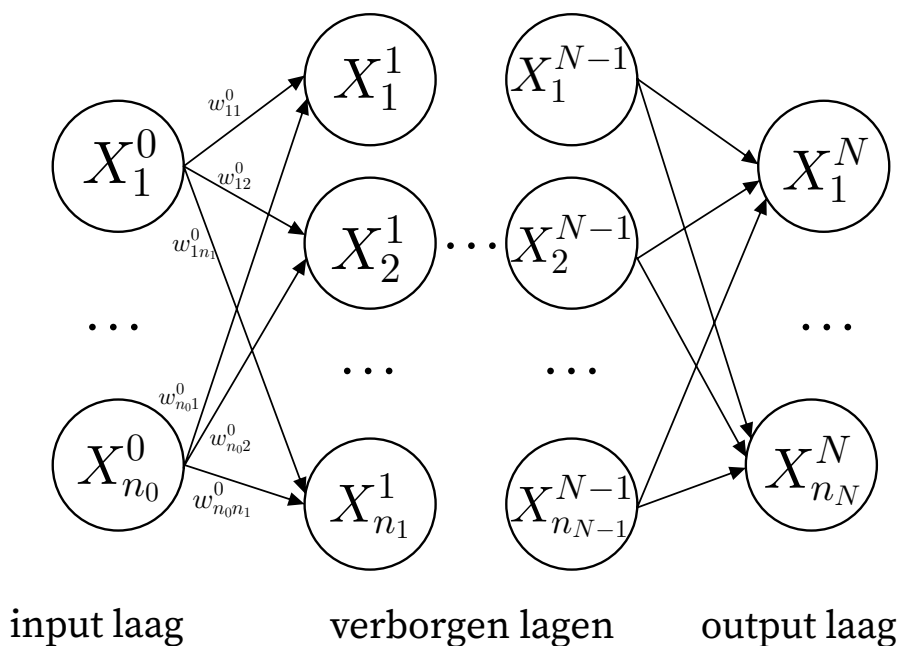
In de huidige generatie van artificiële intelligentie (AI) kunnen data-gedreven methoden een immens gamma aan problemen verwerken. Het is nuttig het verschil tussen AI, machine learning en deep learning te benadrukken. AI is het overkoepelende vakgebied omtrent het bouwen van intelligente systemen. Machine learning is een vakgebied binnen AI waarin een systeem leert a.d.h.v. data i.p.v. expliciet geprogrammeerd te worden. Deep learning is een vakgebied binnen machine learning waarin diepe neurale netwerken (NN) gebruikt worden. Enkele wel-gekende vakgebieden waar deep learning

revolutionaire resultaten boekt in vergelijking met algoritme-gebaseerde en klassieke statistische technieken zijn beeldverwerking en natuurlijke taalverwerking. NN komen in een breed gamma aan variaties voor en vormen de huidige state-of-the-art in zo goed als alle machine learning toepassingen. Bijgevolg winnen deep learning technieken in het laatste decennium erg aan populariteit als hulpmiddel in wetenschappelijk onderzoek.

Een onderscheid moet gemaakt worden tussen gesuperviseerde en ongesuperviseerde machine learning technieken. Een gesuperviseerde techniek steunt op het gebruik van één of meerdere waarheidslabels horend bij elk object in een dataset. Deze labels bevatten informatie over een eigenschap van het object dat door het model voorspeld moet worden. Het model leert door te trachten een kostfunctie – zoals bv. het verschil tussen zijn voorspelling van deze eigenschap en het waarheidslabel – te minimaliseren. Bij ongesuperviseerd leren, zoals bv. principal component analyse, wordt geen gebruik gemaakt van deze waarheidslabels. Er kan dus ook geen kostfunctie geminimaliseerd worden. Eerder tracht het model hier om patronen en structuren te vinden in de dataset.

Een gedetailleerde beschrijving van deep learning en de werking van de vele verschillende soorten NN is buiten de scope van deze thesis. Hiervoor wordt verwezen naar toegewijde boeken zoals bv. van Goodfellow et al. [3] of Aggarwal [4]. Er wordt een beknopt overzicht gegeven van de meest cruciale concepten in deep learning die nodig zijn om wat volgt in deze thesis te begrijpen. Volgend overzicht is gebaseerd op het handboek van Aggarwal [4].

Eén van de meest eenvoudige soorten van NN is een meerlaags perceptron (MLP), zoals voorgesteld in figuur 1.2. Moderne NN architecturen, zoals convolutionele neurale netwerken (CNN), recurrente neurale netwerken en transformers, voegen extra operaties toe aan de werking van een MLP. Deze worden verder besproken in hoofdstuk 3. De fundamenteën van de voorspellingen en training van al deze soorten NN blijven echter hetzelfde. Een voorspelling volgt uit één voorwaartse propagatie van een reeks input-data doorheen het NN. Training van een NN volgt uit een opvolging van voorwaartse en achterwaartse propagatie doorheen het netwerk. Deze concepten kunnen best wiskundig beschreven worden a.d.h.v. een MLP.



Figuur 1.2. Opbouw van een meerlaags perceptron

Een MLP bestaat uit opeenvolging van zgn. volledig verbonden lagen, elk bestaande uit een aantal nodes die een bepaalde activiteit aannemen. Elke node is verbonden met alle nodes uit de voorgaande laag. Een MLP krijgt als input een vector, waarvan element n de activatie X_n^0 vormt van node n van de input laag. De activatie van node i van de volgende laag wordt berekend als een gewogen som van de activaties van alle nodes uit de voorgaande laag. Bij deze gewogen som wordt een node-afhankelijk bias opgeteld, en het totaal wordt door een niet-lineaire functie Φ gehaald.

$$X_i^{j+1} = \Phi \left(\sum_n (w_{ni}^j X_n^j) + b_i^{j+1} \right) \quad (1.1)$$

Dit kan eenvoudig herschreven worden in termen van matrix operaties.

$$\bar{X}^{j+1} = \Phi (W^j \bar{X}^j + \bar{b}^{j+1}) \quad (1.2)$$

De gewichten-matrices W en de bias-vectoren \bar{b} vormen de leerbare parameters van het NN, welke geoptimaliseerd worden doorheen de training. De zgn. activatiefunctie Φ is een te kiezen vaste functie die essentieel is om het NN niet-lineaire data te laten modelleren. Dit proces wordt herhaald doorheen alle lagen van het MLP tot aan de output laag, die een voorspelling maakt over een eigenschap van de input data.

De werking van het MLP is zeer eenvoudig en haalt zijn vermogen tot het leren van

complexe representaties van data volledig uit het grootschalig maken van dit systeem. MLP hebben typisch miljoenen leerbare parameters die geoptimaliseerd worden a.d.h.v. grote hoeveelheden trainingsdata. Zoals eerder vermeld maken NN gebruik van gesuperviseerd leren. Voor elke input vector is er dus een waarheidslabel dat zegt wat de gewenste waarde van de output vector \bar{O} , berekend door voorwaartse propagatie, moet zijn. Er kan dan een kostfunctie $L(\bar{O}, \bar{O}_{\text{waarheid}})$ berekend worden die een maat is voor de afwijking van de output – bij de huidige parameters – tot het waarheidslabel. Vervolgens wordt achterwaartse propagatie gebruikt om de huidige parameters bij te werken en zo de kostfunctie te minimaliseren. Hiertoe wordt de gradiënt van de kostfunctie berekend i.f.v. elke parameter, door de multivariate kettingregel sequentieel van de achterste naar de voorste laag toe te passen. De afgeleide van de kostfunctie naar een gewicht w_{ab}^{r-1} kan dus geschreven worden als volgt, waarbij de node-indices voor de eenvoud weggelaten worden.

$$\frac{\partial L}{\partial w^{r-1}} = \frac{\partial L}{\partial O} \cdot \left[\sum_{\text{Paden } X^r \rightarrow O} \frac{\partial O}{\partial X^k} \prod_{i=r}^{k-1} \frac{\partial X^{i+1}}{\partial X^i} \right] \cdot \frac{\partial X^r}{\partial w^{r-1}} \quad (1.3)$$

De zgn. optimizer bepaalt dan hoeveel de waarde van het gewicht bijgewerkt wordt. Er zijn verschillende soorten optimizers, welke allemaal sterk afhankelijk zijn van de learning-rate hyperparameter¹. De keuzes van optimizer en learning-rate bepalen hoe snel en hoe stabiel het NN door training convergeert naar een (lokaal) minimum van de kostfunctie. In de praktijk worden de parameters van het NN niet bijgewerkt na elke individuele input vector. Eerder worden de gesommeerde gradiënten van een aantal input vectoren gebruikt voor één bijwerking van de parameters. Zo een groep input vectoren wordt een batch genoemd. De keuze van de batchgrootte is ook een hyperparameter die de snelheid en de stabiliteit van het training-proces beïnvloedt.

1.3 Kader van de studie en gerelateerd werk

Grote hoeveelheden EELS-data verworven in bv. een STEM-scan moeten nog steeds manueel geïnterpreteerd worden door een expert. Ook model-gebaseerde kwantificatie vereist eerst een elementidentificatie door een expert. In spectrale beelden wordt vaak slechts een sterk gereduceerde hoeveelheid data bestudeerd voor identificatie, waarna de gevonden elementaire inhoud verondersteld wordt voor alle spectra. Een manuele interpretatie van alle spectra kost de expert te veel tijd. Een data-gedreven techniek kan de taak van elementidentificatie van een expert overnemen en kan wel elk spectrum in een spectraal beeld analyseren. Bovendien is de manuele interpretatie van een expert mogelijks beïnvloed door een neiging naar elementen die verwacht of veelvoorkomend

¹Een hyperparameter is een parameter die het ontwerp van een NN beschrijft, en mag niet worden verward met de leerbare parameters zoals de gewichten en biases. Klassieke voorbeelden van hyperparameters zijn het aantal lagen in een NN of het aantal nodes in een laag.

zijn. Men kan hiertoe beargumenteren dat een expert enkel nuttige voorkennis hanteert bij de classificatie van spectra. Een data-gedreven techniek als tweede opinie in de interpretatie van EELS spectra kan mogelijks de hoeveelheid vals-negatieve identificaties van onverwachte elementen verbeteren.

Vroege pogingen tot geautomatiseerde identificatie en kwantificatie van edges steunde op het gebruik van filters in combinatie met getabelleerde eigenschappen van edges, voornamelijk de onset energie [5][6]. Deze methodes hadden beperkt succes in praktijk-toepassingen omwille van de hoge ruisgevoeligheid en ongevoeligheid voor edges met een lage piek tot achtergrond verhouding. Ongesuperviseerde technieken, zoals bv. K-means clustering, non-negative matrix factorization en autoencoders werden in EELS reeds uitgebreid toegepast voor spectrale decompositie [7]–[10]. Gesuperviseerde technieken, zoals NN en support vector machines, geven mogelijkheid tot meer algemene EELS toepassingen zoals bepaling van oxidatie-staat [11]–[13], zero-loss piek bepaling [14], spectrale deconvolutie [15] en voorspelling van fase transitie [16]. NN werden ook succesvol toegepast in veel gelijkaardige experimentele technieken zoals X-Ray diffractie [17], vibrationele spectroscopie [18][19], laser-induced breakdown spectroscopie [20], X-ray fluorescentie spectroscopie [21], energie dispersieve X-Ray spectroscopie [22] en moleculaire excitatie spectroscopie [23].

Het is opmerkelijk dat deep learning tot recent geen toepassing had gevonden in elementidentificatie in EELS spectra. Dit is te wijten aan – zoals in veel andere toepassingen van deep learning – het gebrek aan bruikbare data. Het verzamelen en labelen van de nodige hoeveelheden data is erg arbeidsintensief. Kong et al. [24] ontwikkelden – tijdens het maken van het hier gepresenteerd werk – een synthetische dataset met de K of L core-loss edge van 20 veelvoorkomende elementen² voor identificatie. Hun simulatiemethode steunt op het verwerken van experimentele data. Core-loss edges in experimentele spectra worden van de spectrum-achtergrond geïsoleerd d.m.v. multi-Gaussian fitting en gevarieerd door deze te schalen, verschuiven en ruis toe te voegen. Meerdere edges en een achtergrond worden dan gecombineerd tot een synthetisch spectrum. De voornaamste limitatie van deze aanpak is de beperkte hoeveelheid experimentele data, in het bijzonder voor zwaardere elementen. In de dataset voorgesteld door Kong et al. zijn er meerdere elementen, zoals bv. S en Cl, waarvoor alle gesimuleerde spectra op slechts één experimentele edge gebaseerd zijn. Hierbij treedt het gevaar op dat het NN slecht generaliseert in zijn voorspelling op nieuwe instanties van deze edge. Het gebruik van grote hoeveelheden trainingsdata met voldoende variatie is fundamenteel voor het succes van deep learning. NN leren uit de trainingsdata representaties van de objecten die invariant zijn voor de natuurlijke variatie van deze objecten. Deze natuurlijke variatie is in dit geval bv. de fijnstructuur van een edge. Een gebrek aan deze natuurlijke variatie binnen de trainingsdata maakt dat de repre-

²Namelijk B, C, N, O, F, Ca, Sc, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Si, P, S, Cl, Ar

sentaties geleerd door het NN willekeurige eigenschappen als absoluut beschouwen. Bijgevolg tredt zgn. overfitting op waarbij verkeerde voorspellingen worden gemaakt op nieuwe instanties van een object. Een tweede limitatie aan de dataset van Kong et al. is het gebruik van één vaste detector range van 50 tot 1100 eV voor alle spectra. Het gevolg hiervan is dat het NN ook enkel voorspellingen kan maken op spectra die dit volledige gebied opmeten, wat de praktijktoepassingen sterk beperkt. Om toch een voorspelling te maken op een spectrum dat niet deze volledige range invult, moeten alle edges geïsoleerd worden van de echte experimentele achtergrond en geplaatst worden op een gesimuleerde achtergrond. Dit proces is kostelijk wanneer het moet gebeuren voor alle spectra in een spectraal beeld en bemoeilijkt een real-time voorspelling aan de microscoop. Bovendien beïnvloedt dit proces ook de integriteit van het spectrum. Edges met een lage piek tot achtergrond verhouding of lage signaal-ruis verhouding kunnen hierbij verloren gaan, ook al zou het NN wel in staat zijn geweest deze te detecteren.

De dataset gevormd door Kong et al. werd gebruikt voor het trainen van een convolutioneel bi-directioneel long short-term memory netwerk voor elementidentificatie. Ze rapporteren dat dit NN 94.9 procent van de spectra in de testset volledig correct classificeert. Hierbij moet echter een belangrijke kanttekening gemaakt worden. De beperking tot 20 elementen in de dataset van Kong et al. brengt een probleem van mogelijke vals positieven mee, bovenop het probleem van vals negatieven. Veronderstel dat een NN, getraind op deze dataset met 20 elementen, een voorspelling moet maken op een spectrum dat een edge bevat van een element dat niet in de dataset voorkomt, zoals bv. een Sr M edge bij 133 eV. Het is dan onvermijdelijk dat het NN een vals negatieve voorspelling maakt voor deze Sr edge. Het NN zal waarschijnlijk wel de aanwezigheid van een edge opmerken, en trachten deze te linken met één van de gekende edges, zoals in dit geval bv. de P L edge bij 132 eV. Een dataset die gelimiteerd is in het aantal gekende elementen is dus vatbaar voor vals negatieven die gekoppeld zijn aan een hoge waarschijnlijkheid voor vals positieven. Hierbij moet opgemerkt worden dat zowel deze vals negatieve en vals positieve voorspellingen niet bijdragen tot de metrieken die het NN van Kong et al. evalueren, omdat de testset ook enkel de gekende elementen bevat. Dit impliceert dat de gerapporteerde 94.9% kans op een juiste evaluatie enkel verwacht kan worden indien geweten is dat de data waarop je het NN toepast enkel deze 20 elementen bevat. Maar deze vooronderstelling is hoogstwaarschijnlijk niet gegrond, daar het NN net wordt toegepast op data waarvan de elementaire inhoud niet gekend is. Deze problematiek benadrukt het belang van een dataset die zo dicht mogelijk aansluit bij het bevatten van alle mogelijke edges die voorkomen in het energieverliesgebied waarop voorspellingen gemaakt worden.

In deze thesis wordt een volledig gesimuleerde dataset voorgesteld, die 107 K, L, M of N edges in 80 elementen representeert. Figuur 1.3 geeft weer welke elementen en respectievelijke edges in de dataset voorkomen. De dataset bevat 736 000 spectra met

Vorming van de dataset

2.1 EELS theorie

In deze sectie wordt een overzicht gegeven van de theorie omtrent elektronenverstrooiing nodig voor de simulatie van EELS spectra, gebaseerd op het handboek van Egerton [25].

2.1.1 Core-loss excitaties

Een core-loss edge wordt beschreven a.d.h.v. de cross-sectie σ voor inelastische verstrooiing van een snel elektron aan een atoom. Het transitie-matrixelement ϵ_n voor inelastische verstrooiing wordt gedefinieerd als

$$\epsilon_n = \int \psi_n^* \sum_j \exp(i\mathbf{q} \cdot \mathbf{r}_j) \psi_0 d\tau = \left\langle \psi_n \left| \sum_j \exp(i\mathbf{q} \cdot \mathbf{r}_j) \right| \psi_0 \right\rangle \quad (2.1)$$

Met ψ_0 de golffunctie van de initiële toestand, ψ_n de golffunctie van de finale toestand, $\hbar\mathbf{q} = \hbar(\mathbf{k}_0 - \mathbf{k}_1)$ de impuls overgedragen aan het atoom – met \mathbf{k}_0 en \mathbf{k}_1 de golfvectoren van het snel elektron respectievelijk voor en na verstrooiing – en \mathbf{r} de positie van het snel elektron. De integratie loopt over alle volume elementen $d\tau$ binnen het atoom. Hiermee wordt de dynamische structuurfactor $|\epsilon_n(q)|^2$ gedefinieerd, die voorkomt in de definitie van de generalized oscillator strength (GOS) $f_n(q)$

$$f_n(q) = \frac{E_n |\epsilon_n(q)|^2}{R (qa_0)^2} \quad (2.2)$$

Met R de Rydberg energie, a_0 de Bohrstraal en E_n de energie geassocieerd met de transitie. Uit de GOS volgen de differentiële cross-sectie $\frac{d\sigma_n}{d\Omega}$ en – in het geval van een continuüm van finale toestanden – de dubbel-differentiële cross-sectie $\frac{d^2\sigma}{d\Omega dE}$

$$\frac{d\sigma_n}{d\Omega} = \frac{4\gamma^2 R k_1}{E_n q^2 k_0} f_n(q) \quad (2.3)$$

$$\frac{d^2\sigma}{d\Omega dE} = \frac{4\gamma^2 R k_1}{E q^2 k_0} \frac{df}{dE}(q, E) \quad (2.4)$$

Met $d\Omega$ de infinitesimale ruimtehoek naar waar verstrooid wordt en γ de Lorentz factor die de relativistische massa van het elektron beschrijft. De energie-differentiële cross-sectie wordt verkregen door vergelijking 2.4 te integreren over de collectiehoek β of over q . Integratie over q , gecombineerd met het toepassen van behoud van momentum, levert

$$\frac{d\sigma}{dE} \approx \frac{4\pi\gamma^2 R}{Ek_0^2} \int \frac{df(q, E)}{dE} \frac{d(q^2)}{q^2} = \frac{4\pi a_0^2 R^2}{ET} \int \frac{df(q, E)}{dE} d[\ln(qa_0)^2] \quad (2.5)$$

Met $T = m_0\nu^2/2$, waarbij m_0 de rustmassa van het elektron is en ν de snelheid van het snel elektron is. De cross-sectie σ_k volgt dan als

$$\sigma_k(\beta, \Delta) = \int_{E_k}^{E_k+\Delta} \frac{d\sigma}{dE} dE \quad (2.6)$$

2.1.2 Low-loss gebied

Het low-loss gebied kan gemodelleerd worden d.m.v. een zero-loss piek en een reeks plasmon pieken.

De zero-loss piek kan goed beschreven worden als een Lorentz piek rond $E_0 = 0$ met een volle breedte bij halve hoogte w

$$ZLP(E) = A \frac{w^2}{w^2 + (2E - 2E_0)^2} = A \frac{w^2}{w^2 + 4E^2} \quad (2.7)$$

De plasmon pieken worden gemodelleerd door het Drude model, waarbij de energie-differentiële cross-sectie gegeven wordt door

$$\frac{d\sigma(\beta)}{dE} \approx \left(\frac{1}{\pi a_0 m_0 \nu^2 n_a} \right) \frac{E(\Delta E_p) E_p^2}{(E^2 - E_p^2)^2 + (E\Delta E_p)^2} \ln \left(1 + \frac{\beta^2}{\theta_E^2} \right) \quad (2.8)$$

Met E_p de plasmon energie, ΔE_p de breedte van de plasmon piek, n_a het aantal atomen per eenheid volume van het medium en $\theta_E = E/(\gamma m_0 \nu^2)$ de karakteristieke hoek.

De zero-loss piek en de plasmon pieken worden onderling geschaald volgens de waarschijnlijkheid P_n dat een elektron n botsingen ondergaat. Onder de veronderstelling dat elke botsing onafhankelijk is, volgt deze waarschijnlijkheid een Poissonproces beschreven door

$$P_n = \frac{1}{n!} \left(\frac{t}{\lambda} \right)^n \exp\left(-\frac{t}{\lambda}\right) \quad (2.9)$$

Waarbij de verstrooiing-parameter $\frac{t}{\lambda}$ het gemiddelde aantal verstrooiingen beschrijft, met t de dikte van het sample en λ de inelastische gemiddelde vrije padlengte.

2.1.3 Fijnstructuur

Vergelijking 2.5 in combinatie met de GOS voor vrije atomen beschrijft de cross-secties voor verstrooiing aan vrije atomen. Zowel het low-loss gebied als de core-loss edges vertonen in realiteit een fijnstructuur in de vorm van pieken of oscillaties in de intensiteit. Deze fijnstructuur resulteert uit de invloed van omringende atomen rond het geïoniseerde atoom, en is dus gekoppeld aan de kristallografische- of energiebanden structuur van het materiaal. De eigenschappen van de fijnstructuur kunnen begrepen worden a.d.h.v. de lokale dichtheid van onbezette toestanden boven het Fermi-niveau. Het lokale karakter van de toestanden maakt dat de fijnstructuur verschillend is voor de edges van verschillende elementen in hetzelfde materiaal. Bovendien zorgen de transitie selectieregels er voor dat de fijnstructuur op verschillende edges van hetzelfde element in hetzelfde materiaal over het algemeen verschillen.

De pieken en oscillaties in de fijnstructuur ondervinden levensduur-verbreding. Een hoog-energetische toestand in de conductieband zal sneller vervallen dan een laag-energetische toestand. De energie van zo een toestand gaat gemiddeld verloren in een levensduur τ_f

$$\tau_f \approx \frac{\lambda_i}{\nu} \quad (2.10)$$

Hierbij is λ_i de inelastische gemiddelde vrije padlengte en $\nu = \sqrt{2\epsilon/m}$ de snelheid van het conductie-elektron, met ϵ de energie boven de edge onset. De energie-verbreding volgt dan uit het onzekerheidsprincipe van Heisenberg

$$\Gamma_f \approx \frac{\hbar}{\tau_f} \quad (2.11)$$

De inelastische gemiddelde vrije padlengte werd geparametriseerd – voor vaste stoffen bestaande uit slechts één element – als volgt [26]

$$\lambda_i = 538a\epsilon^{-2} + 0.41a^{3/2}\epsilon^{1/2} \quad (2.12)$$

Met λ_i en de atomaire diameter a beide uitgedrukt in nanometer. Ten gevolge van de levensduur-verbreding zal de fijnstructuur bestaan uit smallere pieken dicht bij de edge onset en vervallen in bredere oscillaties naarmate de energie toeneemt. Daar de breedte van de piek omgekeerd evenredig is met de amplitude, zal de maximale amplitude die de fijnstructuur kan aannemen groter zijn dicht bij de edge onset. De maximale amplitude van een fijnstructuur piek kan meermaals de maximale amplitude van de core-loss edge voor het vrije atoom zijn. Dergelijke uitgesproken fijnstructuurpieken dicht bij de edge onset noemt men white lines. Figuur 1.1 vertoont een duidelijk voorbeeld van fijnstructuur op zowel de zuurstof K en cerium M edges. De scherp gepiekte intensiteit net achter de onset van de cerium M_5 en M_4 edges is een duidelijk voorbeeld van white lines.

2.2 Synthetische dataset

2.2.1 Inhoud van de dataset

De dataset gesimuleerd door Kong et al. – die gelimiteerd is door de beschikbaarheid van experimentele spectra – bevat de K of L core-loss edge van 20 veelvoorkomende elementen. Een volledig synthetische dataset heeft het voordeel dat deze een veel bredere groep aan elementen en edges toelaat. Figuur 1.3 toont de elementen en bijhorende edges die aanwezig zijn in de dataset. H, He en Li zijn niet betrokken omdat hun K edge zich in het low-loss gebied bevindt. Elementen zwaarder dan Bi worden niet betrokken omdat ze extreem zeldzaam zijn. Op uitzondering van Ac, Th, Pa, U, Np en Pu zijn er geen elementen zwaarder dan Bi die voorkomen op The Materials Project [27]. Ook deze zes uitzondering hebben een zeer lage abundantie en zijn niet van groot belang in EELS.

De abundantie van elk element in de trainingsdata heeft een significante invloed op de voorspellingen van een NN. Het is daarom erg belangrijk doordachte keuzes te maken in de vorming van de samples die gesimuleerd worden. Om een grote variatie aan samples te simuleren kan gekozen worden om willekeurige combinaties van elementen te maken, volgens willekeurige verhoudingen. Deze keuze werd gehanteerd in de dataset gemaakt door Kong et al. In dit geval moet een keuze gemaakt worden omtrent

1. De kansverdeling van het aantal elementen in één sample
2. De kansverdeling van de relatieve verhoudingen tussen de elementen in één sample
3. Het gewicht van elk element in de totale dataset

Voor punt één en twee kan gekozen worden voor een arbitraire maar brede verdeling, die niet gegrond is met de verdeling in de natuur, maar de natuurlijke gevallen wel omvat. Deze aanpak is inefficiënt daar een zeer groot deel van de dataset zal bestaan uit samples die niet realistisch zijn. Een andere aanpak is om punt één en twee zo goed mogelijk te doen samenvallen met de verdeling in de natuur. Het is niet triviaal om een dergelijke kansverdeling te simuleren. Een eenvoudigere methode om een natuurlijke verdeling voor punt één en twee te verkrijgen, is door bestaande samples te kiezen uit de lijst met materialen op The Materials Project. Deze methode wordt gekozen voor de hier gepresenteerde dataset. Binnen deze methode moet nog steeds gekozen worden hoe punt drie te verdelen. Eén extreem is om te trachten alle elementen in de dataset een gelijke aanwezigheid te geven. Het andere extreem is om te trachten zo goed mogelijk de natuurlijke abundantie te behouden, m.a.w. de verhoudingen

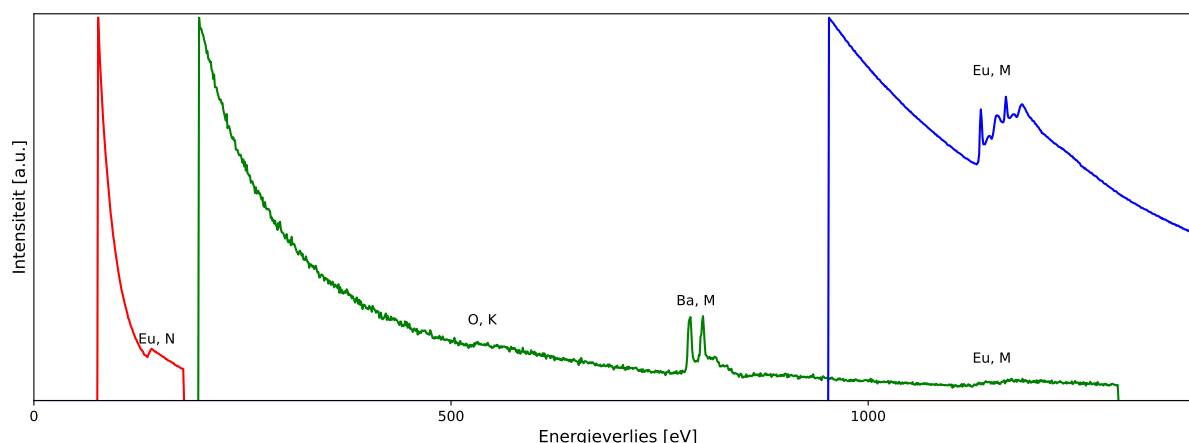
waarin de elementen voorkomen op The Materials Project. Deze keuze reflecteert of de natuurlijke verdeling van elementen beschouwd wordt als nuttige voorkennis, of als bias. Bij een gebrek aan context is ambiguïteit over bepaalde edges onvermijdelijk, daar verschillende edges kunnen voorkomen bij eenzelfde onset energie. Een expert zal, bij gebrek aan voldoende informatie, de identificatie van een edge dan ook voor een deel baseren op hoe frequent bepaalde elementen voorkomen. Het is nuttig deze voorkennis te verwerken in de dataset. Anderzijds is het niet gewenst dat een NN een edge classificeert op basis van veel-voorkomendheid, eerder dan op basis van informatie beschikbaar in het spectrum. Hiertoe wordt getracht een middenweg tussen de twee extremen te volgen. Voor elk element worden evenveel samples gegenereerd waar dit element zeker in zit. De bijkomende elementen, die de rest van het sample vormen, volgen de natuurlijke verdeling van op The Materials Project. Elk sample heeft dus een zgn. query element en mogelijks een reeks van bijkomende elementen. In de lijst van gevormde samples zullen bv. zuurstof en neon evenveel voorkomen als query element. Zuurstof zal echter wel een veel groter gewicht dragen in de totale dataset dan neon, daar zuurstof veelvoorkomend is als bijkomend element en neon niet. Op deze manier worden alle elementen minstens met een gegeven aantal spectra gerepresenteerd in de dataset.

Een bijkomend gevolg van het gebruik van samples uit The Materials Project is dat de dataset inherent informatie bevat over de combinaties waarin elementen voorkomen. Daar de trainingset 112 000 willekeurig gekozen samples bevat, wordt verwacht dat dit slechts een beperkt effect heeft op de voorspellingen. Een voorbeeld waarin wel verwacht kan worden dat dit een uitgesproken effect heeft, is neon. Neon komt in The Materials Project nooit voor als gebonden aan andere elementen. Een NN getraind met deze dataset zal er waarschijnlijk naar neigen om neon opnieuw als alleen voorkomend te voorspellen. Dit kan zowel als positieve voorkennis of negatieve bias beschouwd worden. Een studie van dergelijke randgevallen volgt in sectie 3.3.9.

2.2.2 Werkwijze van de simulatie

De bovenstaande sectie bespreekt de vorming van de samples in de dataset. Deze kunnen beschouwd worden als het analoog van het sample dat in de microscoop wordt geplaatst. Het is hierbij niet gegarandeerd dat alle elementen in dit sample effectief gerepresenteerd zullen zijn in het opgemeten spectrum. Het is mogelijk dat de core-loss edge van een element in het sample niet samenvalt met het opgemeten energieverliesgebied. Net zoals in de microscoop een keuze wordt gemaakt voor het energieverliesgebied dat opgemeten wordt, wordt in de simulatie van een spectrum deze keuze gemaakt. Bovendien zijn enkele elementen in de dataset gerepresenteerd door meerdere edges, maar hoeft er slechts één edge in het spectrum voor te komen om het element te kunnen identificeren. Spectra volgend uit hetzelfde sample kunnen dus een zeer verschillende

elementaire inhoud vertonen. Een voorbeeld hiervan wordt weergegeven in figuur 2.1.

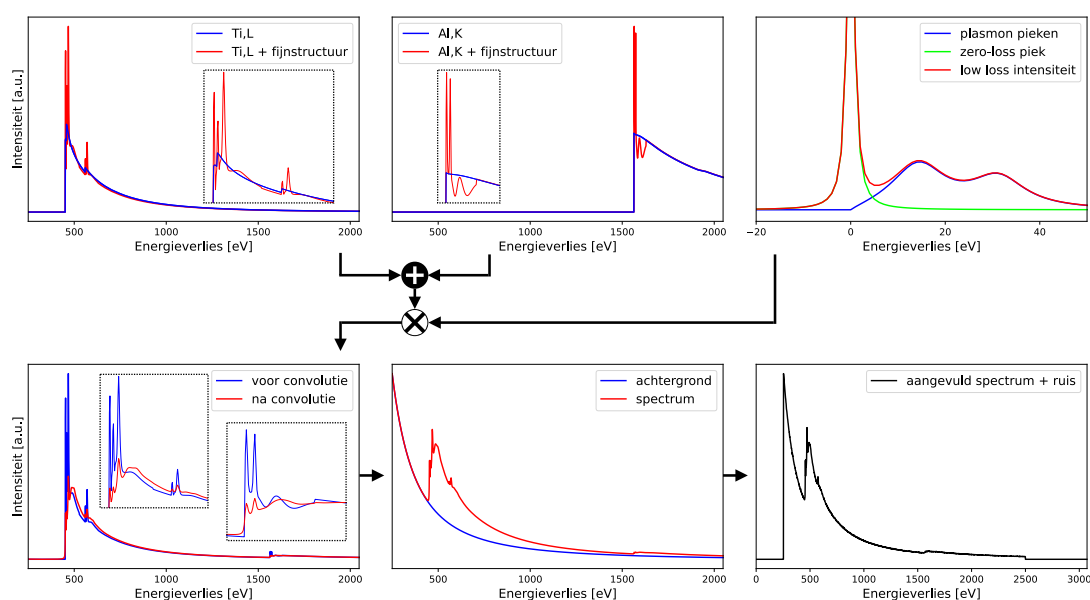


Figuur 2.1. Drie gesimuleerde voorbeeld spectra van hetzelfde sample, namelijk BaEu_2O_4 met Eu als query-element

De elementaire inhoud van een spectrum kan dus verschillen van de elementaire inhoud van het sample, afhankelijk van het willekeurig gekozen energieverliesgebied dat opgemeten wordt. De keuze van het opgemeten energieverliesgebied is echter niet volledig willekeurig, maar wordt onderworpen aan enkele voorwaarden. Het maximaal toegelaten energieverliesgebied gaat van 75 tot en met 3071 eV. Alle spectra worden buiten het effectief opgemeten gebied met intensiteit nul aangevuld in het gebied nul tot en met 3071 eV, zodat elk spectrum dezelfde lengte heeft. Dit is noodzakelijk daar een NN slechts één vaste vorm van input kan ontvangen. Deze nul-padding beperkt sterk de vereisten op het opgemeten energieverliesgebied van experimentele spectra in praktische toepassing. Een tweede voorwaarde die opgelegd wordt is dat het energieverliesgebied wel minstens de major edges van het query element omvat. Deze voorwaarde is noodzakelijk om het voldoende vaak voorkomen van elk query element – wat gegarandeerd werd in de vorming van de samples – te behouden. Indien deze voorwaarde niet opgelegd wordt zouden edges dicht bij de rand van het toegelaten gebied een kleinere kans hebben op het voorkomen in het opgemeten gebied. Bijgevolg zou bv. de Be K edge bij 111 eV nauwelijks voorkomen in een spectrum. Deze voorwaarde zorgt er bovendien voor dat er geen ongecontroleerd groot aantal lege spectra wordt gesimuleerd. Het bevatten van een edge in het opgemeten energieverliesgebied houdt in dat het gebied minstens 10 eV voor de onset van deze edge start, en minstens voor 25 eV achter deze onset verder gaat. Dit impliceert meteen dat de minimale lengte van het energieverliesgebied 35 eV is. In de gevallen waar het query element door twee groepen van edges gerepresenteerd wordt, zoals bv. de M en N edges voor de lanthanoiden, moet deze regel verfijnd worden. Eerst wordt met gelijke kansen gekozen of de eerste groep,

de tweede groep of beide groepen aanwezig moeten zijn in het spectrum. In geval van enkel de eerste of tweede groep kan teruggevallen worden op de standaard regel. In het geval met beide groepen wordt de spectrum start minstens 10 eV voor de onset van laagste-energie major edge gekozen, en het spectrum einde minstens 25 eV achter de onset van de hoogste-energie major edge. Het willekeurig gekozen energieverliesgebied kan gekozen worden tot een energieverlies groter dan het maximale toegelaten energieverlies van 3071 eV. Dit is een bewuste keuze die de kans vergroot dat het opgemeten gebied doorloopt tot het einde van het spectrum, zonder dat zero-padding aan het einde nodig is. De reden hiervoor is dat experimentele spectra vaak een energieverliesgebied opmeten dat voorbij 3071 eV gaat.

Nu de inhoud van het spectrum volledig gespecificeerd is in termen van edges en het opgemeten energieverliesgebied bepaald is, kan een spectrum gesimuleerd worden. Er zijn echter nog tientallen parameters die de uitkomst van de simulatie beïnvloeden. Het doordacht kiezen van de verdelingen waaruit deze parameters gekozen worden is cruciaal voor het vormen van een realistische dataset, en bijgevolg voor het maken van een NN dat zinvolle voorspellingen maakt op experimentele spectra. Tabel 2.1 geeft de waardes weer van alle nodige te kiezen parameters, of de verdelingen waaruit ze willekeurig gekozen worden. Figuur 2.2 geeft een voorstelling van de belangrijkste stappen in de simulatie. Elk onderdeel van dit proces zal in wat volgt in detail besproken worden.



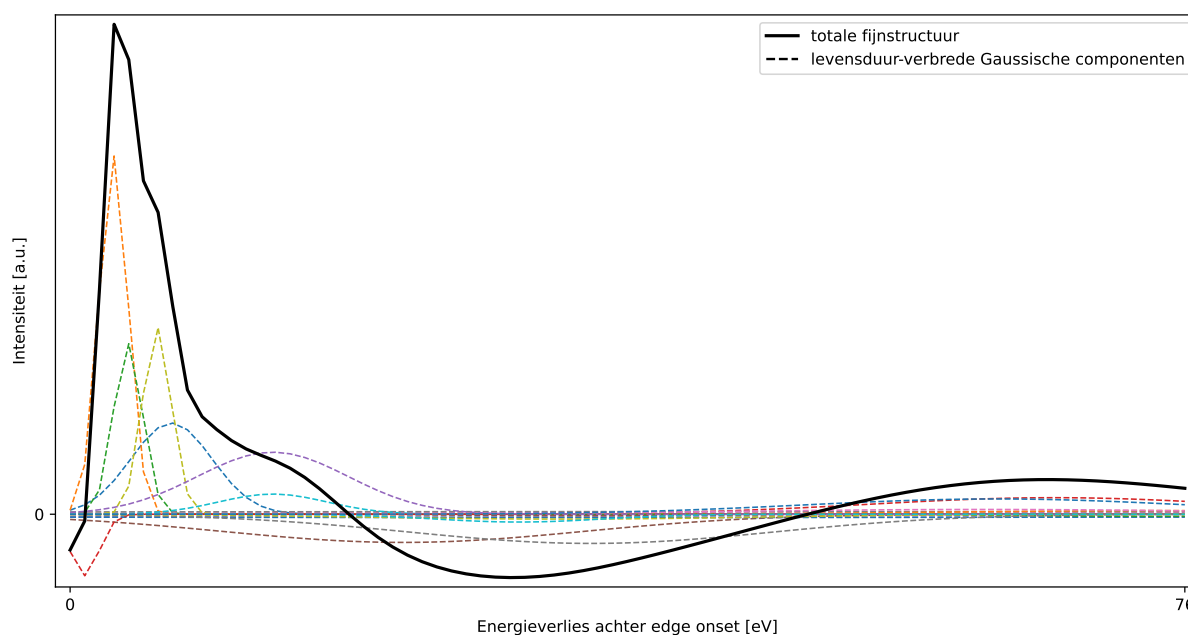
Figuur 2.2. Werkwijze van de simulatie van een TiAl spectrum

Core-loss edge De simulatie van core-loss edges maakt gebruik van publiek beschikbare getabelleerde GOS [28]. De berekende edge onset energieën – welke vatbaar zijn voor afwijkingen – worden vervangen door de getabelleerde onset energieën van EELS.info [1]. Een implementatie van de theorie beschreven in sectie 2.1.1 was reeds beschikbaar in het python pakket pyEELSmodel [29]. Met behulp van dit pakket kunnen de beschikbare GOS tabellen numeriek geïntegreerd worden tot core-loss cross-secties. Hierbij worden zowel de major als de minor edges berekend. De versnelspanning en convergentiehoek van de elektronenbundel en de collectiehoek van de microscoop zijn vereist als parameter. Bij de onset energie van elke edge wordt een chemische verschuiving opgeteld. Deze chemische verschuiving kan het gevolg zijn van verschillende fysische verschijnselen. Een duidelijk voorbeeld van zo een verschijnsel dat een chemische verschuiving veroorzaakt is het bestaan van een band gap. Hierbij wordt de energie van de eerst toegelaten toestand, en dus de edge onset energie, verschoven.

Fijnstructuur Daar de core-loss edges berekend worden a.d.h.v. GOS tabellen voor vrije atomen, bevatten deze niet de materiaalafhankelijke fijnstructuur. Het is niet mogelijk de correcte fijnstructuur te betrekken zonder materiaalafhankelijke berekeningen te doen. Het is computationeel extreem kostelijk om zulke berekening te doen voor elk materiaal in de dataset. Er wordt daarom gekozen te steunen op een willekeurige generieke fijnstructuur die niet tracht de reële fijnstructuur horend bij een gegeven sample te benaderen, maar wel een profiel heeft dat voldoende gelijkenis vertoont met een reële fijnstructuur. Indien de verzameling van alle reële fijnstructuren dan een deelverzameling is van de verzameling met generieke fijnstructuren, kan verwacht worden dat een NN getraind a.d.h.v. generieke fijnstructuren ook goed generaliseert naar reële fijnstructuren. Het effect van fijnstructuur op een core-loss edge wordt gemodelleerd door zulke generieke fijnstructuur – m.a.w. een willekeurige oscillerende functie in een gekozen gebied achter de edge onset – te sommeren op de berekende edge. In model-gebaseerde kwantificatie werd reeds aangetoond dat het sommeren van een puntsgewijs-lineaire functie op de cross-sectie van een vrij atoom geschikt is voor het modelleren van reële fijnstructuur [30]. Om het effect van fijnstructuur te simuleren kan dit idee geïnverteerd worden. Een eerste benadering voor een generieke fijnstructuur kan dus gevormd worden door de intensiteit van elk punt uit een puntsgewijs-lineaire functie willekeurig te kiezen binnen een bepaald bereik. Een dergelijke generieke fijnstructuur zal een erg discontinu karakter vertonen. Een betere benadering wordt bekomen door een curve, zoals bv. een kubische spline, te fitten aan een willekeurige puntsgewijs-lineaire functie. Wanneer de punten van deze puntsgewijs-lineaire functie equidistant gekozen worden, zal de resulterende generieke fijnstructuur geen levensduur verbreding vertonen. Bijgevolg zal die generieke fijnstructuur zelden white lines of het typische gedempte karakter van reële fijnstructuur vertonen.

Om het effect van levensduur-verbreding te betrekken in de generieke fijnstructuur

wordt gekozen om te steunen op een som van willekeurige Gaussische componenten, i.p.v. een willekeurige puntsgewijze functie. De generieke fijnstructuur wordt gevormd uit een gewogen som – met willekeurige gewichten tussen -1 en 1 – van een willekeurig aantal Gaussische pieken die elk voorkomen bij een willekeurige energie en een breedte hebben gelijk aan de levensduur-verbreding bij die energie. Figuur 2.3 geeft een grafische voorstelling van deze methode.



Figuur 2.3. Vorming van een generieke fijnstructuur door gewogen sommatie van een reeks willekeurige levensduur-verbrede Gaussische componenten

De levensduur-verbreding kan berekend worden i.f.v. het energieverlies achter de edge onset d.m.v. vergelijking 2.10-2.12. Een eerste beperking van deze techniek is dat hierbij wordt gesteund op een parametrisatie van de gemiddelde vrije padlengte die bepaald is voor vaste stoffen bestaande uit slechts één element, terwijl deze toegepast wordt op algemene vaste stoffen. Een tweede beperking van deze techniek is dat deze parametrisatie afhankelijk is van de ongekende atomaire diameter. De atomaire diameter wordt daarom willekeurig gekozen tussen 25 en 250 picometer. Er kan dus niet verwacht worden dat de berekende levensduur-verbreding exact zal overeenkomen met de reële levensduur-verbreding in het gegeven sample. Voor toepassingen in het genereren van generieke fijnstructuren is de exacte waarde van de levensduur-verbreding echter niet van groot belang. Het is voldoende om een resultaat te bekomen van de correcte grootteorde met een correcte energieafhankelijkheid.

Voordat de generieke fijnstructuur gecombineerd kan worden met de berekende core-loss edge, moet bepaald worden hoe de amplitudes van beide zich verhouden. In het geval van white lines kan de maximale positieve amplitude van de fijnstructuur meerdere malen groter zijn dan de maximale amplitude van de core-loss edge zonder fijnstructuur. Het sommeren van een generieke fijnstructuur op de berekende core-loss edge mag echter niet in een negatieve cross-sectie resulteren. Om dit te vermijden worden volgende stappen gevolgd:

1. De bekomen generieke fijnstructuur wordt vermenigvuldigd met het teken van het punt met maximale absolute amplitude. Vervolgens wordt de fijnstructuur genormaliseerd. Dit zorgt er voor dat elk punt f van de generieke fijnstructuur zich steeds bevindt binnen het interval $-1 < f \leq 1$.
2. De factor F van de fijnstructuur wordt gedefinieerd als volgt, met f_{\min} de maximale negatieve amplitude, indien deze bestaat.

$$F = \begin{cases} \infty & \text{als } \forall f : f \geq 0 \\ \left| \frac{f_{\max}}{f_{\min}} \right| = \left| \frac{1}{f_{\min}} \right| & \text{als } \exists f < 0 \end{cases}$$

De waarde van F beschrijft hoe veel verder de maximale positieve amplitude van de fijnstructuur reikt dan de maximale negatieve amplitude. Dit vormt een handig, en toch eenvoudig, vertrekpunt voor de verhouding die de fijnstructuur kan aannemen t.o.v. de berekende core-loss edge.

3. De verhouding R – welke de gekozen verhouding is tussen de maximale amplitude van de generieke fijnstructuur en de maximale amplitude van de berekende core-loss edge – wordt dan gekozen als volgt.

$$R \text{ uniform gekozen uit } \begin{cases} [0, F] & \text{als } F \leq R_{\max} \\ [0, R_{\max}] & \text{als } F > R_{\max} \end{cases}$$

De core-loss edge kan nu bijgewerkt worden door hierbij de generieke fijnstructuur, geschaald met R maal de maximale amplitude van de berekende edge, op te tellen. Hierbij moet opgemerkt worden dat de resulterende edge met fijnstructuur later nog geconvolveerd wordt met het low-loss gebied. Hierbij wordt de piek amplitude van de fijnstructuur vaak nog sterk gereduceerd. Een duidelijk voorbeeld hiervan is te zien in figuur 2.2.

4. De bovenstaande stappen garanderen niet dat de resulterende edge met fijnstructuur nooit negatieve waardes aanneemt. Denk bijvoorbeeld aan de situatie waarbij een fijnstructuur met grote negatieve amplitude dicht bij de edge onset opgeteld wordt bij een edge met een verlaat maximum. In dergelijke gevallen worden

alle negatieve amplitudes gelijk gesteld aan nul. Opnieuw zal de convolutie met het low-loss gebied de intensiteit her-distribueren, waardoor er toch een strikt positieve intensiteit bekomen wordt overal in de edge.

5. De volledige core-loss edge met generieke fijnstructuur wordt herschaald zodat zijn integraal gelijk is aan die van de berekende core-loss edge voor toevoeging van fijnstructuur.

Figuur 2.4 geeft een reeks van willekeurige voorbeelden van fijnstructuren die resulteren uit bovenstaande methode.

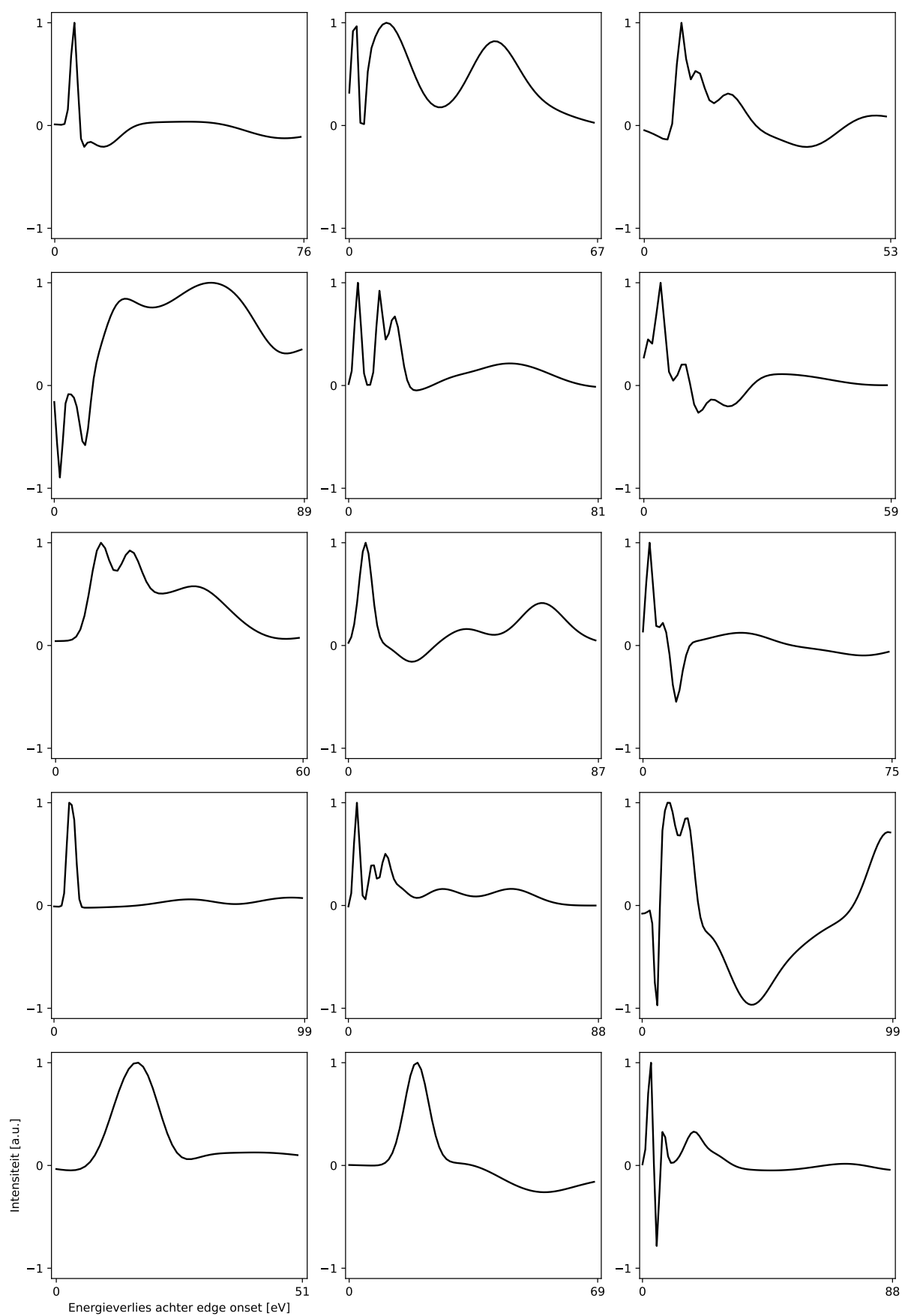
Low-loss gebied Net zoals de fijnstructuur in het core-loss gebied, kan ook het correcte low-loss gebied voor een gegeven sample enkel bekomen worden via kostelijke materiaalafhankelijke berekeningen. Het low-loss gebied is naast de elementaire inhoud en kristallografische structuur van het sample ook afhankelijk van de vorm en dikte van het sample. Opnieuw is de correcte overeenkomst tussen het sample en de berekende intensiteit niet van groot belang en volstaat een generieke benadering. De simulatie van een low-loss gebied, volgens de theorie beschreven in sectie 2.1.2, was reeds beschikbaar in het pyEELSmodel pakket. Een generiek low-loss gebied wordt bekomen door de nodige parameters – zoals bv. de breedte en positie van de plasmon pieken – willekeurig te kiezen binnen een gegeven bereik. Voor de zero-loss piek moet enkel de volle breedte bij halve hoogte (VBHH) gekozen worden. Voor de plasmon excitaties moet het aantal pieken, hun breedte en hun energie gekozen worden. De verstrooiing-parameter moet gekozen worden om de onderlinge verhouding tussen de zero-loss piek en de plasmon pieken te schalen.

Het totale generieke low-loss gebied wordt geconvolveerd over de som van alle core-loss edges. Dit simuleert het effect van meervoudige verstrooiing van een snel elektron bij doorgang van het sample. Een bijkomend voordelig effect van deze convolutie is dat hierbij mogelijke onnatuurlijke discontinuïteiten – zoals degene die kunnen volgen uit het optellen van de fijnstructuur op de berekende core-loss edge – uitgesmeerd worden.

Achtergrond De achtergrond in een core-loss EELS spectrum is het gevolg van de staart van bv. het low-loss gebied of lagere energie core-loss excitaties. De achtergrond is een gladde monotoon dalende functie van het energieverlies, en wordt meestal beschreven met een machtsfunctie van de vorm

$$I_{achtergrond} = A \left(\frac{E}{E_0} \right)^{-r}$$

De achtergrond wordt gesimuleerd volgens deze functie met A en r als te kiezen parameters. De keuze van A beïnvloedt enkel de signaal-ruis verhouding van het spectrum.



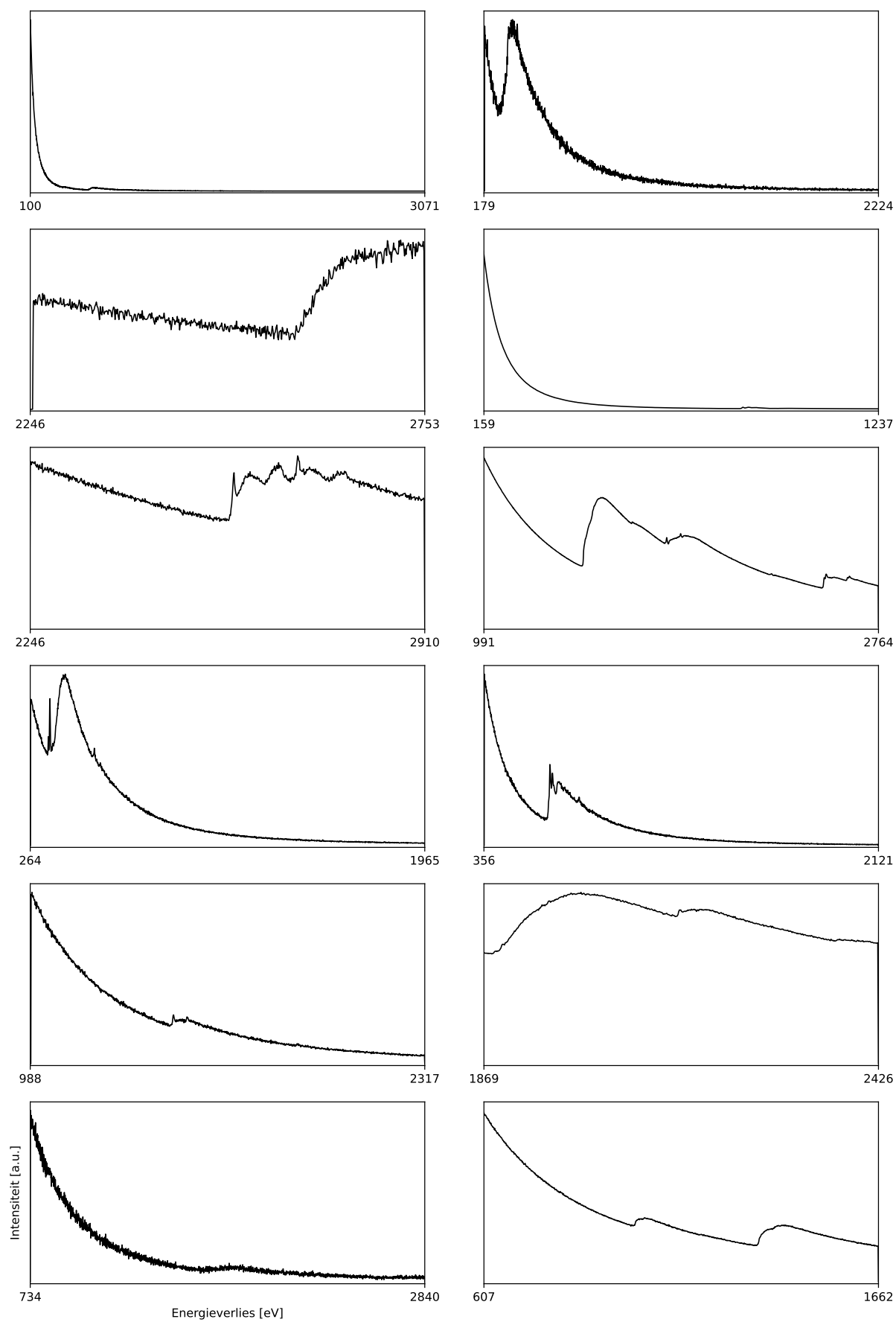
Figuur 2.4. Reeks van willekeurige fijnstructuur voorbeelden

De keuze van r beïnvloedt hoe snel de achtergrond afneemt, en valt voor EELS spectra typisch tussen twee en vier. De jump-ratio wordt gedefinieerd als de verhouding van de maximale intensiteit van de berekende core-loss edge voor fijnstructuur t.o.v. de intensiteit van de achtergrond bij de edge onset. Een jump-ratio, willekeurig gekozen tussen 0.2 en 1.5, wordt opgelegd op één willekeurig gekozen edge in het spectrum. De jump-ratios van de andere edges in het spectrum volgen hierbij automatisch uit de verhoudingen van de cross-secties van elke edge en de intensiteit van de achtergrond bij de onset van elke edge. Deze resulterende jump-ratios kunnen dus ongebonden klein of groot worden.

Bijkomend Elk gesimuleerd spectrum wordt aan Poissonruis onderworpen. De signaalruis verhouding wordt voornamelijk bepaald door de amplitude van de achtergrond. Vervolgens wordt het gehele spectrum onderworpen aan een willekeurige instrumentele verschuiving. In een experimenteel spectrum kan zo een verschuiving optreden als gevolg van bv. een imperfecte uitlijning van de detector en de spectrometer. Vervolgens wordt het spectrum genormaliseerd. Ten slotte wordt het spectrum buiten het opgemeten gebied met nul aangevuld. Deze aanvulling maakt dat alle spectra de juiste lengte hebben om een input te vormen voor een NN. Bovendien worden zo de energie-as informatie en intensiteit informatie verwerkt in één vector, omdat alle spectra hetzelfde energieverlies-nulpunt gebruiken. Figuur 2.5 geeft een reeks voorbeelden van gesimuleerde spectra. De nul-padding wordt hierbij niet getoond om de visualisatie te verduidelijken.

Tabel 2.1. Tabel met de relevante parameters voor de simulatie van een spectrum. Alle waarschijnlijkheidsverdelingen zijn uniform verdeeld.

Parameter	Waarde
Versnelspanning	$\in [60, 100, 200, 300]$ kV
Convergentiehoek	$\in \mathbb{R}$; $\in [1, 20]$ mrad
Collectiehoek	$\in \mathbb{R}$; $\in [1, 100]$ mrad
Energiedispersie	1 eV
Spectrum start	$\in \mathbb{N}$; $\in [75, E_{eerste\ edge} - 10]$ eV
Spectrum einde	$\in \mathbb{N}$; $\in [E_{laatste\ edge} + 25, 3272]$ eV
Totaal spectrum bereik	0 tot en met 3071 eV
VBHH van zero-loss peak	$\in \mathbb{R}$; $\in [1, 3]$ eV
Positie van plasmon pieken	$\in \mathbb{N}$; $\in [3, 20]$ eV
Breedte van plasmon pieken	$\in \mathbb{N}$; $\in [3, 20]$ eV
Aantal plasmon pieken	$\in \mathbb{N}$; $\in [2, 5]$
Verstrooiing-parameter $\frac{t}{\lambda}$	$\in \mathbb{R}$; $\in [0.1, 1]$
Jump-ratio	$\in \mathbb{R}$; $\in [0.2, 1.5]$
Exponent van achtergrond r	$\in \mathbb{R}$; $\in [2.0, 4.0]$
Amplitude van achtergrond A	$\in [10^3, 10^4, 10^5, 10^6, 10^7]$
Breedte van fijnstructuurgebied	$\in \mathbb{N}$; $\in [50, 100]$ eV
Aantal Gauss. comp. in fijnstructuur	$\in \mathbb{N}$; $\in [10, 25]$
R_{max} van fijnstructuur	10
Atomaire diameter	$\in \mathbb{R}$; $\in [25, 250]$ pm
Chemische shift	$\in \mathbb{Z}$; $\in [-5, 5]$ eV
Instrumentele shift	$\in \mathbb{Z}$; $\in [-5, 5]$ eV



Figuur 2.5. Willekeurige voorbeelden van gesimuleerde spectra

2.3 Experimentele dataset

197 experimentele core-loss EELS spectra werden verzameld uit online databanken waarvan 83 bij het Electron Energy Loss Data Center (EELDC) [31], 94 uit de EELS atlas op de EELS.info website van Gatan[1] en 20 spectra bij de Electron Energy-Loss Spectroscopy and X-Ray Absorption Spectroscopy Database (EELS DB) [32]. Enkel de EELS DB vermeldt indien beschikbaar de afkomst van de data, namelijk [30], [33]–[40]. De experimentele dataset bevat in totaal 279 core-loss edges.

Daar een NN slechts één vaste vorm van input kan ontvangen zal het in het algemeen nodig zijn om opgemeten spectra te verwerken vooraleer deze aan het NN aan te bieden. Het is wenselijk om te trachten deze voor-verwerking zo minimaal mogelijk te houden, om een snelle doorgang van opgemeten data naar het NN te garanderen en te streven naar een real-time voorspelling aan de microscoop. Wanneer de energiedispersie van een opgemeten spectrum veel kleiner is dan de vereiste dispersie, in dit geval 1 eV, is het voordelig om het spectrum opnieuw te binnenvoeren. Vervolgens kan, indien nodig, een kubische spline of een andere vorm van interpolatie gebruikt worden het spectrum uit te drukken i.f.v. de correcte energie-as.

Er moet opgemerkt worden dat de abundantie van koolstof in de experimentele spectra zeer onzeker is, daar koolstof vaak aanwezig is in de samplehouder van de microscoop. Bijgevolg is de koolstof K edge regelmatig als spoorelement detecteerbaar in het spectrum, al is het met een zeer lage jump-ratio en signaal-ruis verhouding. Hierbij bestaat geen objectieve regel voor het al dan niet markeren van koolstof als aanwezig. Koolstof komt echter ook voor als element in het sample. Het waarheidslabel van koolstof in de experimentele dataset telt koolstof enkel mee wanneer het onderdeel is van het sample, maar koolstof wordt niet in rekening genomen bij het bepalen van de numerieke metrieken die een NN evalueren.

2.4 Verdeling van training-, validatie- en testset

Om verwarring in wat volgt te voorkomen wordt het verschil tussen een training-, validatie- en testset benadrukt. De trainingset bevat de objecten met waarheidslabels die door het model gebruikt worden om te leren. De trainingset bevat het grootste deel van de gehele dataset. Een typische keuze voor de hoeveelheid data die naar de trainingset gaat ligt tussen 60 en 80 procent. De validatie- en testset bevatten opnieuw objecten met waarheidslabels, en worden gebruikt om het model te evalueren. Het is daarom nodig dat de gehele dataset voor verdeling willekeurig geordend is, zonder overlap. De validatieset wordt gebruikt om het ontwerp van het model en de hyperparameters te optimaliseren, maar niet voor het optimaliseren van de eigenlijke parameters van

het NN door training. De testset wordt slechts eenmaal gebruikt om het uiteindelijke volledig getraind en geoptimaliseerd NN te evalueren.

Het aantal experimentele spectra is beperkt. Er kan besloten worden al de experimentele data te gebruiken in één testset. In hoofdstuk 3 zal blijken dat het toch nuttig is om de experimentele data willekeurig te verdelen in een validatieset met 97 spectra en een testset met 100 spectra. Tabel 2.2 geeft de inhoud van de gesimuleerde training-, validatie- en testset weer. Voor elk van de 80 elementen worden 1400 training samples, 300 validatie samples en 300 test samples opgehaald op The Materials Project. Voor elk sample van de training- en testset worden 5 willekeurige spectra gesimuleerd en voor elk sample van de validatieset worden 2 willekeurige spectra gesimuleerd ¹.

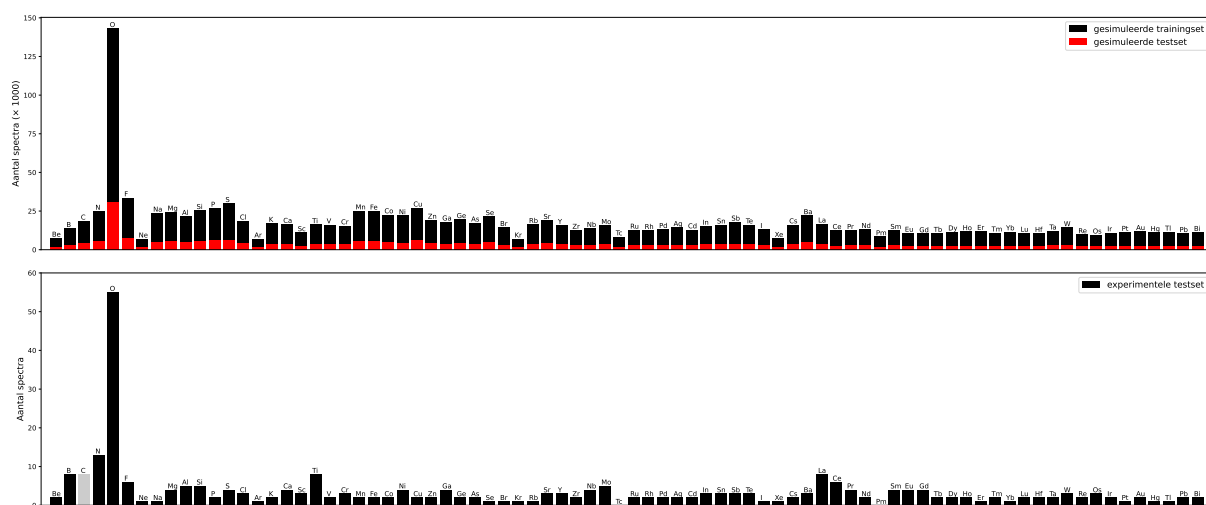
Tabel 2.2. Inhoud van training-, validatie- en testset.

Set	Aantal spectra	Aandeel van totale dataset [%]
Training	560 000	77
Validatie	48 000	7
Test	120 000	16

Bijkomend wordt een dataset gesimuleerd met 8000 spectra – 100 voor elk van de 80 elementen – waarbij elk spectrum slechts edges bevat van één element. Deze dataset zal in sectie 3.3 gebruikt worden om te evalueren welke elementen door een NN met elkaar verward worden.

Ten slotte kunnen de verschillende datasets samengevat worden volgens figuur 2.6, waarbij weergegeven wordt hoe vaak elk element voorkomt in een dataset. De abundantie van elk element in de gesimuleerde dataset is gerelateerd aan de abundanties van elk element op The Materials Project. De abundantie van elk element in de experimentele dataset volgt uit de beschikbaarheid van data op de eerder vermelde databanken.

¹De validatieset is kleiner dan de testset om de kost van een evaluatie bij elke iteratie van het trainingsproces te beperken.



Figuur 2.6. Abundantie van elk element in de gesimuleerde en experimentele datasets. De abundantie van koolstof in de experimentele dataset is onzeker.

Neurale netwerken voor elementidentificatie

3.1 Netwerk architecturen

Wanneer een nieuwe taak – of zelfs een welgekende taak in een nieuwe dataset – wordt aangevangen met NN, is het niet mogelijk a priori te weten welke architectuur het best zal presteren. De conventie is om de meest geschikte architectuur empirisch te bepalen. Hiertoe worden 6 verschillende NN architecturen, waaronder enkele historisch impactvolle en enkele state-of-the-art, geëvalueerd en vergeleken in identificatievermogen en efficiëntie. Tot de vergeleken architecturen behoren: een MLP, een CNN, een U-Net, een Residueel NN, een vision transformer en een compacte convolutionele transformer.

Het is algemeen aanvaard dat een eenvoudig MLP zelden nog competitie vormt voor modernere architecturen, zoals bv. een CNN voor beeldverwerking of een transformer voor natuurlijke taalverwerking. De superioriteit van CNN in beeldverwerking is onder meer te danken aan het verbeterd vermogen tot het leren van translatie-invariantie. In EELS spectra is er slechts een beperkte vereiste tot het aanleren van translatie-invariantie¹. Sterker nog, de onset energie is de voornaamste vorm van informatie. Het is daarom interessant toch een MLP te evalueren. Hierbij moet benadrukt worden dat er nog vele andere redenen zijn waarom het MLP onderdoet voor de huidige state-of-the-art, zoals bv. het grote aantal parameters nodig om een zekere leercapaciteit te bereiken.

CNN zijn bij uitstek de meest gebruikte architectuur in beeldverwerking, en werden populair door het succes van AlexNet [41] en later o.a. VGGNet [42]. CNN bestaan uit een reeks convolutionele lagen, meestal gevolgd door een klein MLP. Een convolutionele laag bestaat uit een aantal filters met een vaste grootte, die over de output van de voorgaande laag worden geconvolueerd. Via achterwaartse propagatie worden de parameters van de filters bijgewerkt i.p.v. de verbindingen tussen verschillende lagen. De uitkomst van de convoluties worden de kenmerkkaarten genoemd. De convolutionele lagen en volledig verbonden lagen lenen zich respectievelijk tot extractie van kenmerken en regressie van deze kenmerken. De vroegste convolutionele lagen hebben een beperkt receptief veld en leren lokale primitieve kenmerken. Diepere convolutionele lagen vormen complexere representaties uit deze primitieve kenmerken. De volledig verbonden lagen kunnen dan een taakspecifieke output leveren door regressie

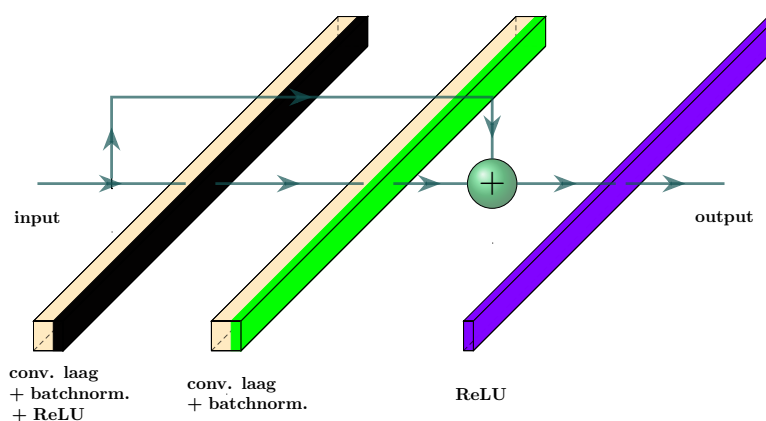
¹enkel in de chemische en instrumentele verschuiving

van de kenmerken geleverd door de convolutionele lagen. Veel CNN bevatten nog een andere belangrijke component, namelijk pooling lagen. Deze aggregeren de output van een convolutie – meestal door het nemen van het gemiddelde of het maximum – en reduceren zo de afmetingen van de bekomen kenmerkkaarten. Het is veelvoorkomend om per dimensie twee naburige punten uit de kenmerkkaart te aggregeren en zo de afmeting van de kaart te halveren. Ook globale pooling is veelgebruikt, waarbij één of meerdere gehele dimensies geaggregeerd worden, zoals bijvoorbeeld alle ruimtelijke dimensies of de kenmerk-dimensie. Globale pooling wordt vooral gebruikt als transitie laag om meerdimensionale kenmerkkaarten om te vormen naar 1D vectoren die aan een MLP kunnen worden gevoed. Deze vormen daarbij een alternatief voor de zgn. flattening laag, die een kenmerkkaart uitrolt tot één – meestal extreem grote – 1D vector. Hierbij wordt een immens aantal parameters geïntroduceerd tussen de flattening laag en het MLP. Het introduceren van nieuwe soorten lagen – zoals convolutionele lagen en pooling lagen – zorgt voor een snelle toename van het aantal mogelijke configuraties. Bovendien hebben deze lagen ook een groot aantal te kiezen hyperparameters, zoals bv. de filter-afmeting, het aantal filters, de pooling afmeting en de pooling richting. Het is daarom veelvoorkomend om een beproefde architectuur te volgen i.p.v. kostelijke experimenten uit te voeren om een goede configuratie te bepalen. Er wordt gekozen om een CNN te gebruiken dat a.h.w. een 1D versie is van het bekende VGG-11 beeldverwerkingsmodel.

De U-Net architectuur [43] is een vorm van CNN waarbij de kenmerkkaarten eerst een ruimtelijk-krimpand en vervolgens ruimtelijk-expanderend pad volgen, welke verbonden worden door voorwaartse kopieer-verbindingen. Deze voorwaartse kopieer-verbindingen stapelen de kenmerkkaarten van de verbonden lagen op. Bijgevolg leent het ontwerp zich sterk tot het gebruik van zowel globale als lokale contextuele informatie. De U-Net architectuur werd geïntroduceerd als beeld-segmentatie netwerk, waarbij de output dus opnieuw een beeld is. Om een classificatie model uit een klassiek U-Net te vormen, worden de convolutionele lagen opgevolgd door een MLP, net zoals in andere CNN.

Residuele neurale netwerken (ResNets) [44] zijn een vorm van CNN ontworpen om het vanishing gradient probleem tegen te gaan, dat ontstaat bij het trainen van steeds diepere NN. Het vanishing gradient probleem houdt in dat de gradiënten volgend uit achterwaartse propagatie steeds kleiner worden naarmate de vroegere lagen van het NN bereikt worden. Daar de incrementele verbetering van een laag evenredig is met deze gradiënten, ondervinden de vroege lagen slechts een zeer beperkte training. He et al. observeerden dat diepere NN slechter presteerden dan zeer gelijkaardige minder diepe varianten [44]. Het diepere NN kan echter minstens dezelfde prestatie behalen als de minder diepe variant, door simpelweg de identiteit-functie te genereren in de extra lagen. Het vanishing gradient probleem maakt dat diepere NN moeilijker te trainen zijn,

m.a.w. minder eenvoudig een voldoende lokaal minimum bereiken. Het introduceren van de zgn. voorwaartse skip-verbinding verbetert de propagatie van gradiënten doorheen het netwerk. Een voorwaartse skip-verbinding sommeert de kenmerkaarten van de verbonden lagen. ResNets maken gebruik van residuele blokken – zoals voorgesteld² in figuur 3.1 – welke bestaan uit twee convolutionele lagen met batchnormalisatie, ReLU activatie en een voorwaartse skip-verbinding.



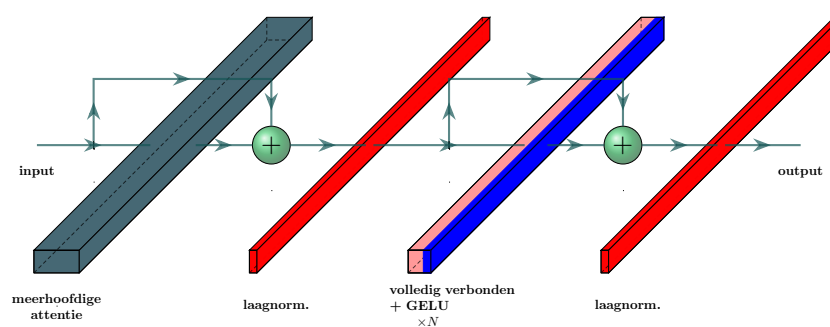
Figuur 3.1. Residuele blok uit een ResNet

Zowel de klassieke CNN, het U-Net en het ResNet uit de laatste paragrafen werden geïntroduceerd voor beeldverwerking. Kong et al. beargumenteerden dat een EELS spectrum zowel als 1D beeld en als sequentiële reeks geïnterpreteerd kan worden. Bijgevolg werd de CNN-biLSTM architectuur door Kong et al. voorgesteld voor elementidentificatie in EELS spectra. Deze architectuur bevat de voornaamste componenten van zowel beeldverwerking- en tijdreeksmodellen. Net zoals CNN een typische keuze zijn voor beeldverwerking, zijn recurrente neurale netwerken (RNN) dit voor reeks-vormige data. In RNN is de activatie van een gegeven node afhankelijk van alle voorgaande nodes in dezelfde laag. Bijgevolg kunnen dit soort netwerken enkel serieel geoptimaliseerd worden. Dit is erg nadelig t.o.v. alle andere architecturen hier voorgesteld, die door parallellisatie op een GPU significant sneller trainen en voorspellingen maken³. De long short-term memory (LSTM) [46] is de meest succesvolle vorm van RNN. De LSTM heeft een periode van groot succes gekend, maar is niet langer de state-of-the-art voor de meerderheid van de tijdreeks en sequentie toepassingen. De huidige state-of-the-art bevindt zich voornamelijk bij architecturen die het attentiemechanisme hanteren, waarvan de meest succesvolle de zgn. transformers zijn [47]. Een attentielag is in

²Alle visualisaties van NN architecturen zijn gebaseerd op de PlotNeuralNet software [45].

³Om hier een concreet voorbeeld van te geven: Een U-Net met 20 mln. parameters kan in minder dan 2 minuten evenveel data verwerken als een CNN-biLSTM met minder dan 300 000 parameters in 20 minuten verwerkt.

staat meer belang te hechten aan een beperkt gebied van zijn input en is bovendien wel volledig parallelliseerbaar. De bekendste vorm van attentiel laag is de zgn. geschaalde dot-product attentiel laag. Een gedetailleerde beschrijving van de werking van het attentiemechanisme is buiten de scope van deze thesis. Het volstaat om te weten dat het attentiemechanisme gebruik maakt van leerbare vectorrepresentaties waarvoor de waarde van hun dot-product toeneemt naarmate de gerepresenteerde objecten meer onderlinge relevantie hebben. Een transformer is een NN gebaseerd op een opvolging van meerdere transformer encoder blokken. Eén transformer encoder blok – zoals voorgesteld in figuur 3.2 – bestaat uit een meerhoofdige attentiel laag, een MLP met GELU activatie, voorwaartse skip-verbindingen en laagnormalisaties. Een meerhoofdige attentiel laag bestaat op zijn beurt uit een reeks van parallele geschaalde dot-product attentielagen.



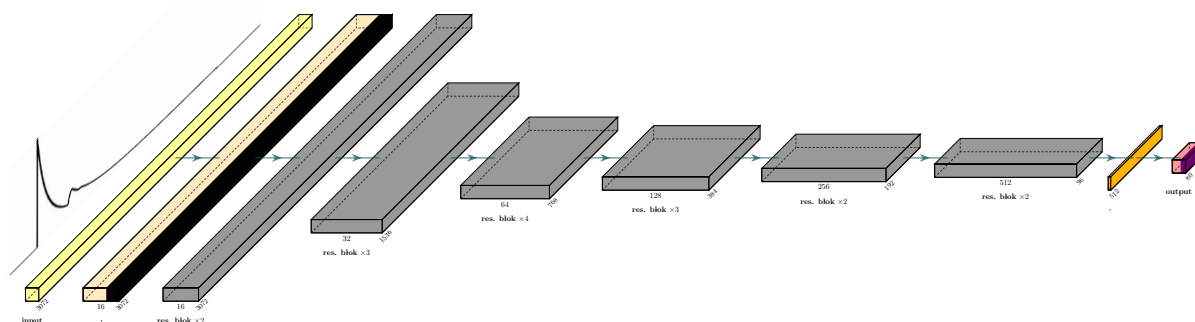
Figuur 3.2. Transformer encoder blok

De transformer werd geïntroduceerd als model voor natuurlijke taalverwerking, waar het immens succes kent in de zgn. generative pre-trained transformers of GPT's [48]. Ook in beeldverwerking kennen transformers groot succes met o.a. de vision transformer (ViT) [49]. In de ViT wordt een afbeelding opgedeeld in een reeks van rechthoekige segmenten, die gecodeerd worden tot vectoren en daarna aan een transformer encoder worden gevoed. De compacte convolutionele transformer (CCT) [50] werd geïntroduceerd om de efficiëntie van de ViT te verbeteren. De CCT gebruikt convolutionele lagen i.p.v. opdeling in segmenten om de input data te coderen tot geschikte vector representaties. De CCT vertoont een zeer grote gelijkheid met de CNN-biLSTM, waar a.h.w. de seriële LSTM vervangen is door een parallelliseerbare transformer.

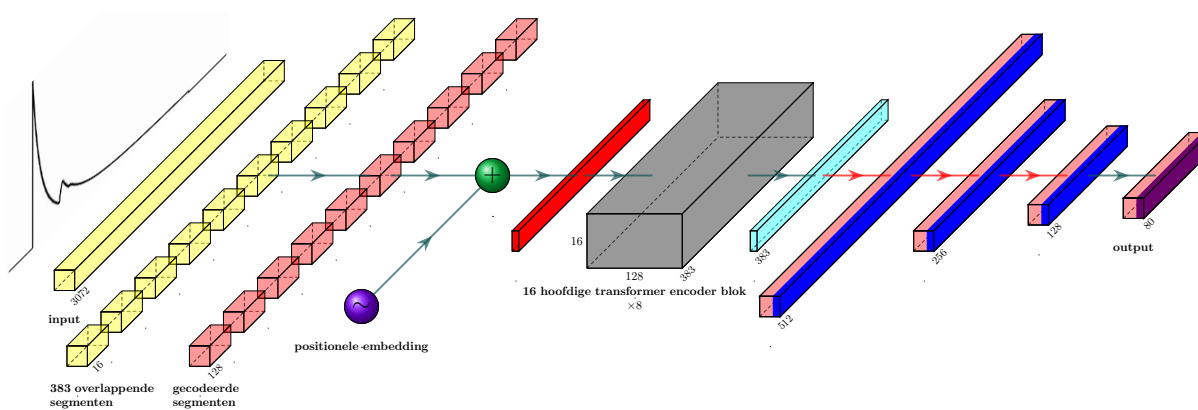
De originele ResNet en ViT architecturen – welke ontworpen zijn voor beeldverwerking – gebruiken global average pooling (GAP) over de ruimtelijke dimensies. GAP vormt hier meerdimensionale kenmerkkaarten om tot 1D vectoren die de input van het MLP vormen. In deze ruimtelijke GAP – in wat volgt vermeld als simpelweg GAP – wordt de ruimtelijke informatie sterk onderdrukt. Elementidentificatie in EELS steunt, in tegenstelling tot veel toepassingen in beeldclassificatie, zeer sterk op deze ruimtelijke

informatie. Bijgevolg wordt verwacht dat het gebruik van GAP inefficiënt is voor deze taak. Ook om een MLP aan de U-Net architectuur toe te voegen is een dimensiereductie-methode nodig. De originele U-Net publicatie gebruikt een leerbare gewogen som over alle kenmerken horend bij één ruimtelijk punt. Deze methode wordt in wat volgt kenmerk weging genoemd. Deze techniek behoudt wel goed de ruimtelijke informatie. Een andere veelgebruikte keuze is de eerder vermelde flattening laag. Deze heeft het voordeel dat alle informatie behouden blijft, in ruil voor een erg grote parameter kost. De ViT en CCT hebben elk nog een eigen unieke methode voor dimensiereductie. De ViT wordt naast GAP ook gebruikt met zgn. tokenization. De CCT introduceert een nieuwe techniek genaamd sequentie pooling, wat overeenkomt met een vorm van leerbare gewogen GAP. De beste dimensiereductie-methode wordt voor elke architectuur empirisch bepaald en de resultaten worden in sectie 3.3.1 besproken. Figuur 3.3 geeft een schematische voorstelling van de vergeleken NN architecturen, waarbij steeds de variant met de beste dimensiereductie-methode wordt weergegeven.

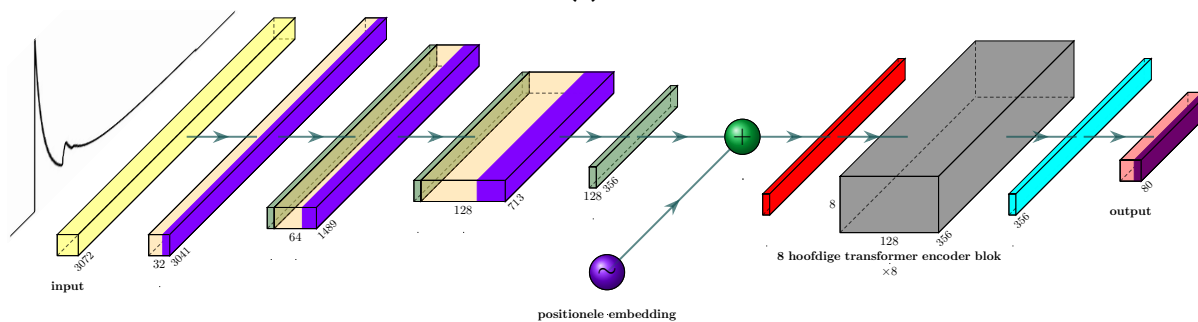
De activatiefuncties worden behouden zoals voorgesteld in de originele publicatie van elke architectuur, op uitzondering van de activatiefunctie van de output laag. Aangezien elementidentificatie een taak is waarbij meerdere klassen toegekend kunnen worden aan één input object, kan geen gebruik gemaakt worden van de veelgebruikte softmax activatiefunctie. Er wordt gekozen voor een sigmoïdefunctie als activatiefunctie in alle output lagen. Deze functie genereert een output tussen nul en één, zodat de voorspelling van een model geïnterpreteerd kan worden als de waarschijnlijkheid dat een element aanwezig is in het input spectrum. Verdere regularisatiemethodes zoals batchnormalisatie en dropout worden gebruikt waar weergegeven in figuur 3.3. De filter-afmeting in convolutionele lagen is steeds 32.



(d) ResNet



(e) ViT



(f) CCT

Figuur 3.3. Visualisaties van de vergeleken NN

3.2 Training en optimalisatie

Elke architectuur wordt getraind met de adaptive moment estimation (Adam) optimizer. Voor elk model wordt de beste learning-rate uit $[10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}]$ bepaald a.d.h.v.

de validatieset. De learning-rate wordt tijdens de training gehalveerd op plateaus om zo de convergentie te verbeteren. De gebruikte batchgrootte is 32. Alle NN worden getraind in TensorFlow 2.11 op een Nvidia GeForce RTX 3060 GPU⁴ met CUDA 11.7 en CuDNN 8.6.

De accuraatheid is de meest gekende metriek voor het evalueren van classificatieproblemen, en wordt gedefinieerd als volgt.

$$\text{Accuraatheid} = \frac{TP + TN}{TP + FP + FN + FP}$$

Waarbij TP staat voor echt positieven, FP voor vals positieven, TN voor echt negatieven en FN voor vals negatieven. In het geval van elementidentificatie wordt deze metriek echter volledig gedomineerd door de grote hoeveelheid echt negatieven, waardoor deze metriek ongeschikt is voor evaluatie. Een betere keuze van metriek voor het evalueren van een classificatieprobleem in een niet-evenwichtig verdeelde dataset is de F_1 -score. De F_1 -score is het harmonisch gemiddelde van de precisie en sensitiviteit. Precisie is gedefinieerd als de fractie van de positieve identificaties die effectief correct is. Sensitiviteit is gedefinieerd als de fractie van de echt positieven die correct geïdentificeerd zijn. In formulevorm kunnen precisie, sensitiviteit en F_1 -score als volgt uitgedrukt worden

$$\text{Precisie} = \frac{TP}{TP + FP} \quad \text{Sensitiviteit} = \frac{TP}{TP + FN} \quad F_1 = \frac{2}{\text{Sensitiviteit}^{-1} + \text{Precisie}^{-1}}$$

Precisie, sensitiviteit en F_1 -score worden berekend per klasse – m.a.w. per chemisch element – en gewogen uitgemiddeld volgens het aantal keer dat elk element voorkomt. Deze gewogen uitmiddeling volgens het onevenwichtige voorkomen van elk element maakt dat de gemiddelde F_1 -score zich mogelijk niet bevindt tussen de gemiddelde precisie en sensitiviteit.

Een bijkomende metriek voor het evalueren van een NN is het percentage van spectra waarvoor de inhoud volledig correct voorspeld wordt, zonder vals positieven of vals negatieven. Dit wordt in wat volgt de exact match ratio (EMR) genoemd.

Alle NN worden getraind met een kostfunctie die de som is van de veelvoorkomende binaire cross-entropie kostfunctie en een zgn. soft F_1 kostfunctie, m.a.w. $L(\bar{O}, \bar{O}_{\text{waarheid}}) = L_{BCE} + L_{F_1}$ waarbij

$$L_{BCE}(\bar{O}, \bar{O}_{\text{waarheid}}) = -\frac{1}{N} \sum_i [y_i \ln(p_i) + (1 - y_i) \ln(1 - p_i)] \quad (3.1)$$

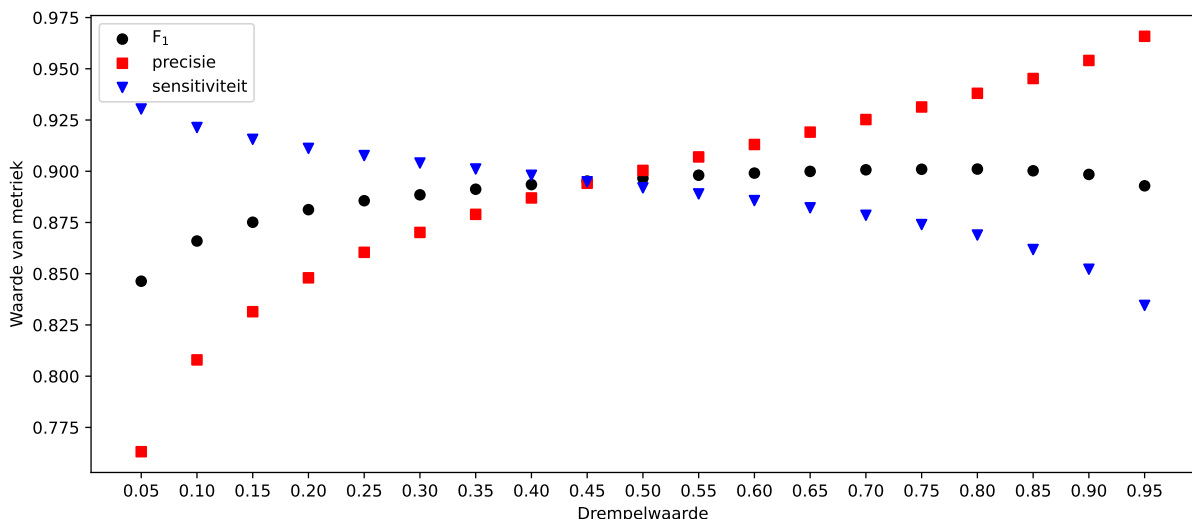
⁴Dit is een standaard GPU. Er bestaan clouddiensten zoals bv. Google Colaboratory en Kaggle die gratis toegang bieden tot GPU's van minstens dit kaliber. Dit maakt dat alle berekeningen die in wat volgt voorgesteld worden zeker toegankelijk zijn voor gebruikers.

$$L_{F_1}(\bar{O}, \bar{O}_{\text{waarheid}}) = 1 - F_{1,\text{soft}} = \frac{1}{N} \sum_i \left[1 - \frac{2\tilde{T}P_i}{2\tilde{T}P_i + \tilde{F}N_i + \tilde{F}N_i} \right] \quad (3.2)$$

Met

$$\begin{cases} \tilde{T}P_i = y_i \cdot p_i \\ \tilde{F}P_i = (1 - y_i) \cdot p_i \\ \tilde{F}N_i = y_i \cdot (1 - p_i) \end{cases}$$

Hierbij is p_i de voorspelde waarschijnlijkheid dat element i aanwezig is in het spectrum en y_i het binaire waarheidslabel dat zegt of element i aanwezig is. Merk op dat deze kostfunctie nergens een drempelwaarde bevat die bepaalt of een voorspelling als aanwezig of afwezig beschouwd moet worden. Het is niet wenselijk om een heavyside-stapfunctie met een gegeven drempelwaarde in de kostfunctie of een activatiefunctie te verwerken, daar deze niet differentieerbaar is. De optimale drempelwaarde moet na de training a.d.h.v. de validatieset bepaald worden. Er werd empirisch bepaald dat het gebruik van deze gecombineerde kostfunctie zorgt voor een betere convergentie van het NN. De optimale drempelwaarde hangt af van de toepassing en voorkeur van de gebruiker. Indien de toepassing meer baat heeft bij een laag aantal vals positieve voorspellingen – m.a.w. een hoge precisie – dan heeft een hoge drempelwaarde de voorkeur. Een toepassing die meer baat heeft bij een laag aantal vals negatieve voorspellingen – m.a.w. een hoge sensitiviteit – prefereert een lage drempelwaarde. In wat volgt wordt gekozen die drempelwaarde te gebruiken waarvoor het verschil tussen precisie en sensitiviteit minimaal is. Deze wordt geschat door de precisie en sensitiviteit voor de validatieset te bepalen bij een reeks van drempelwaarden zoals voorgesteld in figuur 3.4 voor het CNN, waarbij de resulterende drempelwaarde 0.45 is.



Figuur 3.4. Relatie tussen drempelwaarde en precisie, sensitiviteit en F₁-score van het CNN voor de gesimuleerde validatieset

3.3 Resultaten

Voordat de resultaten omtrent de getrainde NN besproken worden, is het belangrijk op te merken dat NN starten van een willekeurige initialisatie. Bovendien wordt de trainingset willekeurig geordend tijdens het trainingsproces. Het is bijgevolg niet deterministisch bepaald in welk (lokaal) minimum een NN na training eindigt. Een herhaling van het trainingsproces kan dus resultaten leveren die meetbaar verschillen van de volgende resultaten. Het is computationeel erg kostelijk om het trainingsproces vele malen te herhalen om deze variaties te kwantificeren. Enkel van het U-Net, de ViT en de CCT worden drie instanties getraind. De resultaten die in wat volgt besproken worden zijn steeds die voor de eerste instantie. De resultaten van de tweede en derde instanties van deze architecturen worden gegeven in appendix A.2. Verder wordt nogmaals benadrukt dat het waarheidslabel van C in de experimentele spectra niet objectief bepaald kan worden, omdat de C K edge zich regelmatig manifesteert vanuit de samplehouder in de microscoop. De voorspellingen voor C worden bijgevolg niet mee in rekening gebracht bij het bepalen van de numerieke metrieken voor experimentele data.

3.3.1 Dimensiereductie-methode

Tabel 3.1 geeft de resultaten weer van alle vergeleken dimensiereductie-methodes⁵. In tegenstelling tot wat intuïtief verwacht kan worden, is de beste methode niet altijd de methode die de ruimtelijke dimensies behoudt. Een mogelijke verklaring voor de observatie dat enkel in het ResNet een dimensiereductie-methode die ruimtelijke informatie lijkt te onderdrukken het beste presteert, bevindt zich in de afmeting van dit netwerk. Het ResNet heeft significant meer diepte en parameters dan de andere vergeleken architecturen. Het is mogelijk dat deze afmeting toelaat representaties te vormen die ruimtelijke informatie goed behouden na globale pooling over de ruimtelijke dimensies. In wat volgt zal voor elke architectuur de dimensiereductie-methode gebruikt worden die in deze evaluatie het beste presteerde.

⁵Voor iedere dimensiereductie-methode werd opnieuw de optimale learning rate bepaald.

Tabel 3.1. Vergelijking van dimensiereductie-methodes. De beste methode wordt steeds in het vet aangeduid.

Model	Methode	Leerbare param. [mln.]	Sim. validatie F ₁ -score
U-Net	GAP	11	0.79
	Kenmerk weging	20	0.86
ResNet	GAP	41	0.89
	Kenmerk weging	41	0.86
	Flatten	45	0.80
ViT	GAP	2	0.52
	Kenmerk weging	2	0.84
	Flatten	27	0.82
	Token	2	0.64
CCT	seq. pool.	5	0.82
	Kenmerk weging	5	0.87

3.3.2 Vergelijking van architecturen

In tabel 3.2 worden de zes verschillende NN architecturen kwantitatief vergeleken i.f.v. hun efficiëntie en prestatie op de gesimuleerde testset en de experimentele dataset met alle 197 experimentele spectra.

Tabel 3.2. Vergelijking van identificatie architecturen

Model	Leerbare param. [mln.]	Inferentie-tijd (CPU GPU) [s]	Sim. test F ₁ -score	Exp. data F ₁ -score	Sim. test EMR	Exp. data EMR
MLP	56	1.0 0.1	0.50	0.45	0.05	0.04
CNN	45	12.7 0.2	0.90	0.72	0.68	0.31
U-Net	20	2.7 0.2	0.86	0.77	0.63	0.53
ResNet	41	8.6 0.3	0.89	0.75	0.68	0.39
ViT	2	19.4 0.6	0.84	0.79	0.55	0.50
CCT	5	20.6 0.6	0.87	0.78	0.60	0.45

Het MLP presteert – niet echt onverwacht – minder goed dan andere modernere en meer geavanceerde architecturen. Het CNN presteert goed op de gesimuleerde data, maar generaliseert in vergelijking met andere architecturen niet goed naar de experimentele data. Het U-Net slaagt er in efficiëntie en prestatie zeer goed te combineren en generaliseert goed naar experimentele data. Ook het ResNet scoort hoog op de gesimuleerde data maar generaliseert minder goed naar experimentele data dan andere architecturen. De ViT behaalt in dit experiment de beste F₁-score op de experimentele

data met een verschil t.o.v. de gesimuleerde testset van slechts 5 procent. Ook de CCT generaliseert goed naar de experimentele data.

De F_1 -score voor de gesimuleerde testset weergegeven in tabel 3.2 is die bij de drempelwaarde waarvoor precisie en sensitiviteit in de gesimuleerde validatieset gelijk zijn. Een verdere studie van de afhankelijkheid van precisie, sensitiviteit en F_1 -score t.o.v. de drempelwaarde toont aan dat de optimale drempelwaarde voor de experimentele spectra verschilt van die voor gesimuleerde spectra. De optimale drempelwaarde voor experimentele spectra is steeds hoger dan die voor gesimuleerde spectra. Dit heeft een sterke invloed op de EMR, wat zich manifesteert in de grote spreiding op de EMR in tabel 3.2. Een waarschijnlijke verklaring voor de nood aan een hogere drempelwaarde voor de experimentele dataset t.o.v. de gesimuleerde datasets bevindt zich in de signaalsterktes in beide. De gesimuleerde datasets bevatten edges met ongebonden-lage jump-ratios⁶ en daaruit volgende zeer slechte signaal-ruis verhoudingen. Een edge met een zeer lage jump-ratio of een zeer lage signaal-ruis verhouding kan gekoppeld worden aan een lage zelfzekerheid in de voorspelling van het NN en bijgevolg aan de nood voor een lage drempelwaarde. Elke edge in de experimentele dataset heeft echter een jump-ratio en signaal-ruis verhouding die de edge duidelijk detecteerbaar maakt.

Om het verschil in drempelwaardes te kunnen corrigeren wordt gekozen de experimentele testset willekeurig te verdelen in een experimentele validatieset met 97 spectra en een experimentele testset met 100 spectra. De resultaten voor de experimentele testset behaald bij de optimale drempelwaarde van de experimentele validatieset worden weergegeven in tabel 3.3. Daar het aantal voorspellingen waarop deze scores gebaseerd zijn nu relatief klein is, is ook de statistische rigorositeit beperkt. Een andere willekeurige verdeling van validatie- en testset kan de behaalde scores significant laten verschillen. Zo zal bv. de F_1 -score van de ViT zakken van 0.83 tot 0.79 wanneer alle experimentele spectra beschouwd worden. Enkel het U-Net en de ViT slagen er steeds in zowel een F_1 -score > 0.8 en EMR ≥ 0.6 te behalen op de experimentele testset (zie ook appendix A.2).

⁶Zoals eerder vermeld wordt een jump-ratio opgelegd op één edge is elk gesimuleerd spectrum en kunnen de jump-ratio's op de andere edges in het spectrum hierbij ongebonden klein worden.

Tabel 3.3. Vergelijking van identificatie architecturen i.f.v. experimentele testdata bij gebruik van de drempelwaarde bepaald i.f.v. experimentele validatiedata

Model	Exp. test F₁-score	Exp. test EMR	Drempel- waarde
MLP	0.42	0.12	0.80
CNN	0.76	0.42	0.95
U-Net	0.81	0.63	0.90
ResNet	0.77	0.53	0.95
ViT	0.83	0.62	0.85
CCT	0.79	0.58	0.85

3.3.3 Graad van overfitting

Het is interessant om de prestatie op de trainingset te vergelijken met de prestatie op de testset om zo te evalueren of een netwerk overfit. De trainingset bevat voldoende data zodat deze slechts éénmaal geïtereerd moet worden om de training van een NN volledig te laten convergeren. Dit beperkt zeer sterk de kans op overfitting. Tabel 3.4 geeft de prestatie weer van de ViT op de gesimuleerde training-, validatie- en testset. De testset bevat samples die opgehaald zijn op The Materials Project onafhankelijk van de samples in de trainingset. Toch is de prestatie op de testset quasi gelijk aan die op de trainingset. Dit toont aan dat het NN niet enkel combinaties van elementen voorspelt zoals ze voorkomen in de trainingset. Verder blijkt dat de F₁-score op de experimentele spectra slechts 5 procent verschilt van die voor de gesimuleerde test set. De EMR is zelfs hoger op de experimentele dataset dan op de gesimuleerde datasets. Zoals eerder vermeld moet hierbij rekening gehouden worden met het idee dat de signaalsterktes in de experimentele dataset beter zijn dan deze in gesimuleerde datasets.

Tabel 3.4. Vergelijking van training-, validatie- en testset en experimentele dataset

Data	F₁-score	Precisie	Sensitiviteit	EMR
Train	0.85	0.85	0.84	0.55
Validatie	0.84	0.85	0.84	0.54
Test	0.84	0.85	0.84	0.55
Exp. data (alle spectra)	0.79	0.82	0.82	0.61

3.3.4 Ensemble van neurale netwerken

Een gekende techniek om de correctheid van een reeks voorspellingen te verbeteren in ruil voor een langere inferentie tijd is d.m.v. ensembles van NN [51]. Hierbij worden

de voorspellingen van een reeks van NN gecombineerd door bv. de voorspelde waarschijnlijkheden per klasse uit te middelen of door elk model een binaire stem uit te laten brengen over de aanwezigheid van elke klasse. De NN die het ensemble vormen kunnen zowel nieuwe instanties zijn van dezelfde architectuur – getraind tot in een ander lokaal minimum van de kostfunctie – of instanties van verschillende architecturen. In wat volgt wordt een ensemble geëvalueerd dat bestaat uit één U-Net en één ViT, een ensemble dat bestaat uit drie U-Net's en een ensemble dat bestaat uit drie U-Net's en twee ViT's. Al deze combinaties van NN kunnen eenvoudig gelijktijdig voorspellingen maken binnen het werkgeheugen van de GPU en bijgevolg in inferentie als een enkel NN behandeld worden⁷. De voorspellingen worden gecombineerd door het nemen van het rekenkundig gemiddelde. Tabel 3.5 toont de behaalde scores op de experimentele testset bij de optimale drempelwaarde van de experimentele validatieset.

Tabel 3.5. Evaluatie van NN ensembles

Opbouw	F₁-score	Precisie	Sensitiviteit	EMR	Drempelwaarde
1×ViT+1×U-Net	0.85	0.87	0.88	0.66	0.75
3×U-Net	0.85	0.87	0.86	0.69	0.85
2×ViT+3×U-Net	0.86	0.88	0.88	0.70	0.75

Het is lastig om deze cijfers te interpreteren omdat slecht in beperkte mate gelijkaardige cijfers geweten zijn voor vergelijkbare EELS-identificatiemethodes. Het is aannemelijk dat een manuele identificatie door een expert hoger zou scoren, maar ook een expert kan edges bij gelijke onset energieën verwarren of edges met lage signaal-ruis verhoudingen over het hoofd zien. Een eerste indruk van deze resultaten zou ook kunnen zijn dat deze minder goed zijn dan de resultaten gerapporteerd door Kong et al. Hierbij moet opgemerkt worden dat geen van bovenstaande resultaten rechtstreeks kunnen worden vergeleken met die gerapporteerd door Kong et al. De taak van het NN hier is beduidend complexer dan die beschreven door Kong et al., omwille van bv. de uitbreiding van het aantal voorspelde labels van 20 naar 80, de uitbreiding en vrijheid van het toegelaten energieverlies gebied, het gebruik van meerdere edges om hetzelfde element te representeren etc.

De behaalde EMR van 70 procent lijkt op het eerste zicht eerder ongunstig. Hier moet opgemerkt worden dat EMR een zeer veeleisende metriek is. Een NN met een EMR van 70 procent mag zeker niet geïnterpreteerd worden als een NN dat slecht in 70 procent van de gevallen nuttige voorspellingen levert. Een voorspelling die voor 79 van de 80 elementen met zeer goed vertrouwen een correcte keuze maakt en slechts voor één element verkeerd uitkomt t.o.v. de drempelwaarde, wordt sterk bestraft in de EMR.

⁷De reden dan geen ensemble van 3 ViT's wordt gebruikt is omdat deze niet gelijktijdig binnen het werkgeheugen van de GPU kunnen rekenen.

Een duidelijk voorbeeld hiervan wordt later besproken in de sectie met kwalitatieve evaluatie.

De F_1 score is moeilijk intuïtief te interpreteren, maar precisie en sensitiviteit hebben duidelijke implicaties. Gegeven de drempelwaarde van 0.75, heeft een element met een core-loss edge aanwezig in het spectrum een kans van 88 procent om effectief gedetecteerd te worden door het $2 \times \text{ViT} + 3 \times \text{U-Net}$ ensemble. Tegelijkertijd is er een kans van 88 procent dat elk element dat als aanwezig voorspeld wordt door het $2 \times \text{ViT} + 3 \times \text{U-Net}$ ensemble, ook effectief aanwezig is. Er kan ook gebruik gemaakt worden van een drempelwaarde die niet streeft naar een gelijke precisie en sensitiviteit. Een drempelwaarde van 0.60 levert een kans van 91 procent dat een aanwezig element effectief gedetecteerd wordt en een kans van 85 procent dat een gedetecteerd element effectief aanwezig is. Een drempelwaarde van 0.90 levert een kans van 81 procent dat een aanwezig element effectief gedetecteerd wordt en een kans van 92 procent dat een gedetecteerd element effectief aanwezig is.

3.3.5 Verwarringsmatrix

Een verwarringsmatrix is een veelgebruikte techniek om te evalueren welke klassen door een NN met elkaar verward worden. Deze verwarring is eenduidig bepaald in toepassingen waar aan elk input object slechts één klasse gekoppeld is. In het geval van elementidentificatie in EELS – waarbij meerdere klassen horen bij één input object – is deze verwarring niet goed gedefinieerd. Hiertoe wordt gebruik gemaakt van een gesimuleerde dataset waarbij elk spectrum slechts één element bevat. Het voorspelde element is dan het element waar het NN de hoogste waarschijnlijkheid aan toekent. Figuur 3.5 geeft de resulterende verwarringsmatrix van het $2 \times \text{ViT} + 3 \times \text{U-Net}$ ensemble.

gevallen als F voorspeld wordt. Het is opmerkelijk dat deze verwarring niet symmetrisch is. Zo worden Sc en F nauwelijks verward voor respectievelijk N en Xe. De asymmetrie van deze verwarring is niet volledig te verklaren door de abundanties van elementen in de dataset, daar N in deze meer voorkomt dan Sc. Verder toont de verwarringsmatrix binnen de groep van lanthaniden een verhoogd aantal verwarringen met directe burenen. Dit is het gevolg van de kleine verschillen in onset energieën tussen de N edges van naburige lanthaniden.

3.3.6 Uitsluiting van N edges

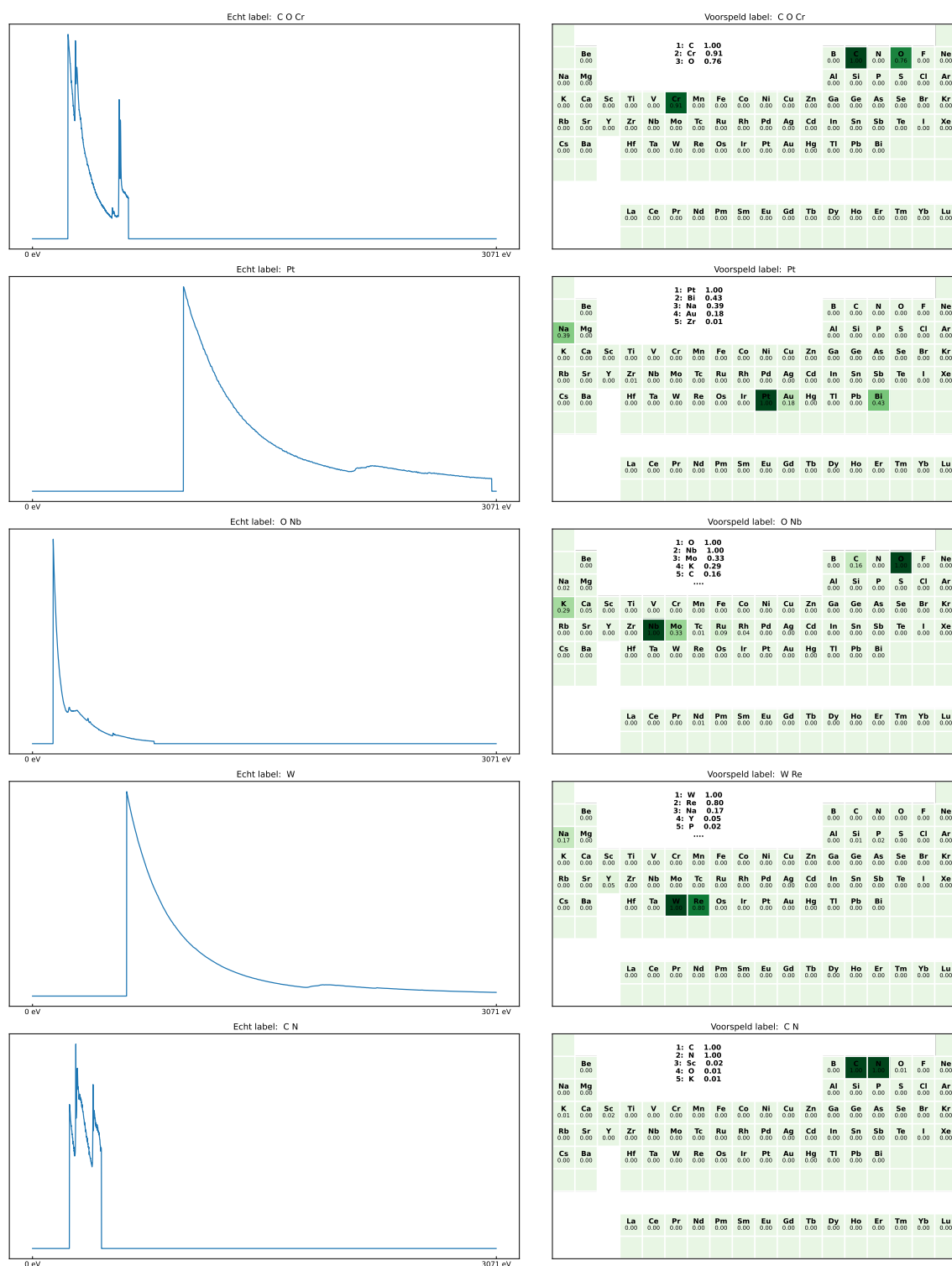
De N edges die voorkomen voor de lanthaniden hebben een zeer klein onderling verschil in onset energie. Het verschil in onset energie van de N edges van twee naburige lanthaniden is vaak kleiner dan de spreiding op de onset energie van 10 eV die gesimuleerd werd. Dit maakt de correcte identificatie van deze bij gebrek aan extra informatie – ook voor een menselijke expert – zeer moeilijk. Tabel 3.6 toont een evaluatie van enkele NN voor de volledige experimentele dataset, uitgezonderd alle spectra met N edges. Hierbij worden 174 experimentele spectra overgehouden. Zoals verwacht zijn de prestaties hier steeds beter dan die voor de dataset met N edges.

Tabel 3.6. Vergelijking van identificatie architecturen op experimentele data uitgezonderd spectra met N edges

Model	F ₁ -score	Precisie	Sensitiviteit	EMR
U-Net	0.86	0.86	0.89	0.70
ResNet	0.81	0.82	0.84	0.60
ViT	0.85	0.85	0.88	0.69
CCT	0.85	0.84	0.89	0.65
2×ViT+3×U-Net ensemble	0.89	0.89	0.92	0.75

3.3.7 Kwalitatieve evaluatie

Een NN kan eenvoudig kwalitatief geëvalueerd worden door de voorspellingen voor enkele voorbeelden te visualiseren. Figuur 3.6 geeft een schematische weergave van enkele voorspellingen van het 2×ViT+3×U-Net ensemble op willekeurige experimentele spectra. Nog 30 extra voorspelling voor willekeurige experimentele spectra worden weergegeven in appendix A.3.

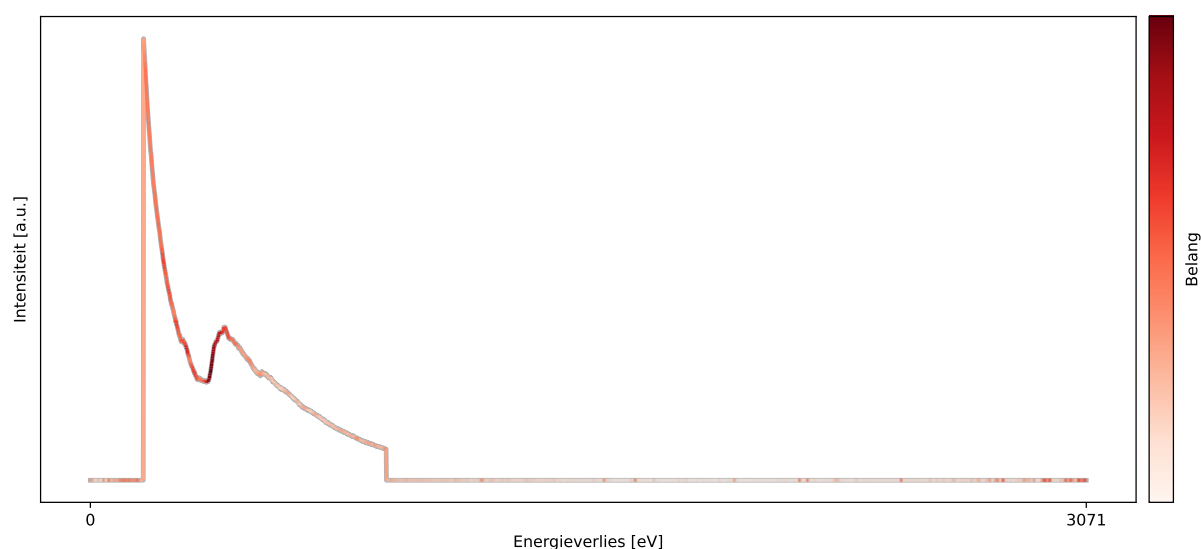


Figuur 3.6. Schematische weergave van de voorspelling van het $2 \times \text{ViT} + 3 \times \text{U-Net}$ ensemble op vijf willekeurige experimentele spectra. Links wordt het input spectrum met waarheidslabel weergegeven. Rechts wordt rechtstreeks de output van het NN weergegeven a.d.h.v. de kleurwaarde en cijfers horend bij elk element. Verder wordt bij elke voorspelling de hoogst scorende elementen weergegeven en het label voorspeld bij gebruik van een drempelwaarde van 0.75.

Het vierde voorbeeld – waarbij in een W spectrum zowel W als Re als aanwezig voorspeld worden – toont een goed voorbeeld van het soort voorspellingen dat zeer streng bestraft wordt in de EMR. Het model slaagt er in correct W met 100 procent zekerheid te identificeren. De zelfzekerheid over de aanwezigheid van Re is 20 procent lager, maar valt wel boven de gekozen drempelwaarde. Wat de EMR betreft is de voorspelling op dit spectrum dus verkeerd, toch zou deze voorspelling zeker nuttige informatie leveren in een praktische toepassing.

3.3.8 Grad-CAM

Inzicht verwerven in de relevante informatie waarop een NN een bepaalde voorspelling maakt is één van de volhardende moeilijkheden in deep learning. Verschillende technieken werden hiertoe reeds voorgesteld. Gradient-based class activation maps of grad-CAM [52] is een veelgebruikt en relatief eenvoudig voorbeeld van een dergelijke techniek. Hierbij wordt de gradiënt van de voorspelling van het NN naar de activatie doorheen een gekozen laag gebruikt als maat voor hoe belangrijk een gebied van de input is. Het belang van elk gebied uit een gekozen laag wordt vervolgens geprojecteerd op de input van het NN. Grad-CAM laat toe het belang van elk gebied binnen één laag te waarden i.f.v. van elk element in de output van het NN. Het belang binnen elke laag van het NN kan ook uitgemiddeld worden over een aantal gekozen lagen. Figuur 3.7 toont het gemiddelde belang van alle meerhoofdige attentielagen in de ViT, geprojecteerd op het input spectrum. Hierbij wordt zoals intuïtief verwacht het gebied rond de edge onset weergegeven als het belangrijkste gebied in de voorspelling.



Figuur 3.7. Gemiddelde Grad-CAM van alle meerhoofdige attentielagen in de ViT voor een spectrum met een Pd M edge

3.3.9 Randgevallen

In deze sectie worden de voorspellingen van het $2 \times \text{ViT} + 3 \times \text{U-Net}$ ensemble op enkele bijzondere gevallen bestudeerd. Eerder werd beweerd dat de vorming van de gesimuleerde samples d.m.v. The Materials Project maakt dat het NN inherent kennis heeft over de natuurlijke combinaties waarin elementen voorkomen. Er werd reeds beargumenteerd dat dit niet resulteert in een NN dat enkel combinaties voorspelt die voorkomen in de trainingset. Aan de hand van Ne kan worden aangetoond dat het NN in bijzondere gevallen wel degelijk steunt op kennis over de combinaties waarin elementen voorkomen. In de samples uit de trainingset die de verdeling op The Materials project volgen, komt Ne nooit voor als gebonden aan een ander element. Op gesimuleerde spectra waarin zowel de Ne K edge als een andere edge voorkomen, steunt het NN wel op de voorkennis uit de verdeling van de dataset en verwacht het Ne met een ander element, voornamelijk Ni. De methode voor het vormen van de samples is er dus in geslaagd een evenwicht te vinden tussen het verliezen van alle voorkennis over de abundantie van elementen en het volledig biasen van het NN.

Een bijkomend experiment onderzoekt of het $2 \times \text{ViT} + 3 \times \text{U-Net}$ ensemble ook in staat is om correct spectra te classificeren waarin geen enkele core-loss edge aanwezig is. Geen enkele van de eerder beschreven datasets bevat dergelijke lege spectra. Tijdens het trainen van het NN is er dus nooit een situatie voorgekomen waarin alle elementen als afwezig voorspeld moeten worden. Indien het NN een goed begrip heeft van wat een core-loss edge voorstelt zou het er toch in moeten slagen lege spectra correct te classificeren. Hiertoe worden voorspellingen gemaakt op 1000 gesimuleerde spectra waarin geen edges aanwezig zijn. Het NN behaalt hier een EMR van 0.86 bij een lage drempelwaarde van 0.35 en een EMR van 1.00 bij een drempelwaarde van 0.75. Dit toont aan dat het NN er in slaagt lege spectra te herkennen ook al komen deze niet voor in de trainingsdata.

Toepassing: Automatische element-mapping

In dit hoofdstuk worden de NN uit het vorige hoofdstuk gebruikt voor een demonstratie van volledig automatische verwerking van EELS spectrale beelden. Een spectraalbeeld heeft voor elke gemeten ruimtelijke coördinaat – in dit geval elke probe-positie van een STEM scan – een corresponderend spectrum. Een identificatie van de elementen aanwezig in elk EELS spectrum resulteert bijgevolg in een mapping van de ruimtelijke verdeling van elementen in het sample. Een manuele identificatie van elementen in elk individueel spectrum vereist extreem veel tijd. Model-gebaseerde kwantificatie laat toe om berekende core-loss edges te fitten aan een gegeven spectrum. Deze techniek laat dus een automatische elementaire mapping toe, indien globaal geweten is welke elementen aanwezig zijn in het sample. Indien echter ook niet geweten is wat de globale inhoud van het sample is, moet deze eerst op manuele of geautomatiseerde wijze bepaald worden¹. Er zijn verschillende methodes mogelijk om op manuele wijze de globale inhoud te bepalen. Een eenvoudige methode bestaat uit het aggregeren van naburige spectra om zo de ruimtelijke resolutie – en bijgevolg het aantal spectra dat geïnspecteerd moet worden – sterk te reduceren. Deze methode heeft de beperking dat elementen met een zeer geringe aanwezigheid dreigen verloren te gaan in de ruis op de achtergrond van de andere spectra. Een meer geavanceerde techniek voor het reduceren van het aantal spectra dat geïnspecteerd moet worden maakt gebruik van ongesuperviseerde machine learning technieken voor spectrale decompositie zoals principal component analyse, K-means clustering en non-negative matrix factorization. Hierbij worden gelijkaardige spectra gegroepeerd, zodat voor elke groep slechts één spectrum manueel geïnspecteerd moet worden. Momenteel is de enige methode om de inhoud volledig automatisch te bepalen de geïntroduceerde deep learning techniek. Dit wordt in wat volgt gedemonstreerd voor een spectraal beeld van $\text{LaMnO}_3/\text{BaTiO}_3/\text{SrTiO}_3$ superrooster en een spectraal beeld van een Ti-O-N sample.

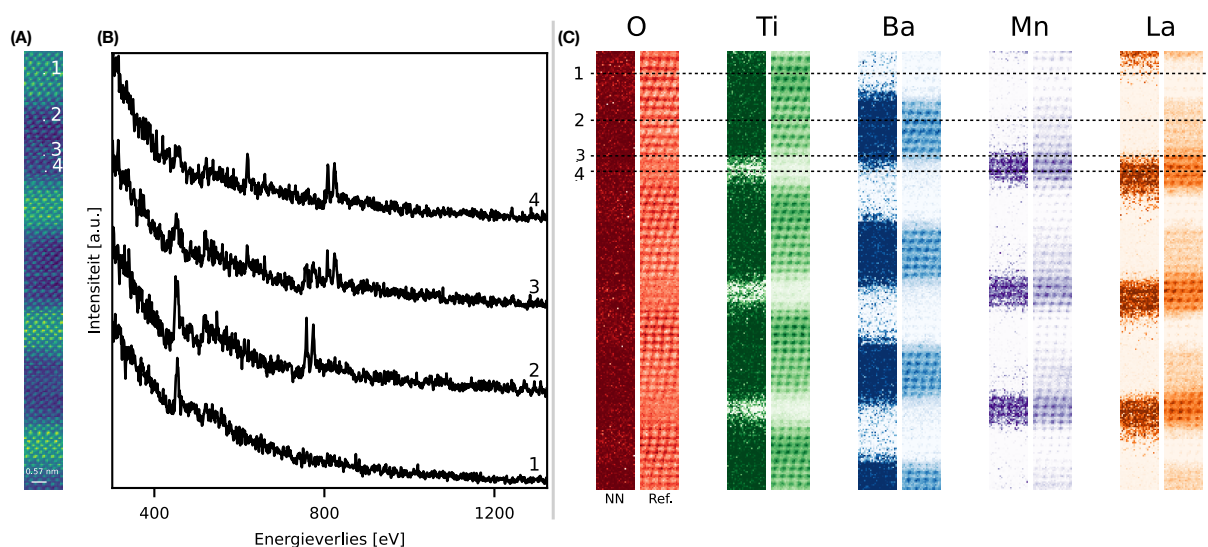
In de verwerking van spectrale beelden worden voorspellingen gemaakt op een groot aantal, vaak eerder ruizige, individuele spectra. Hieruit volgen regelmatig vals positieve voorspellingen die de aanwezigheid van elementen in extreem lage concentraties voorspellen. Hiertoe kan een selectiecriteria worden vooropgesteld om vermoedelijk vals positieve voorspelling te negeren. Het vooropgestelde selectiecriteria is dat enkel

¹In principe kan ook getracht worden alle elementen te fitten aan het spectrum. Deze methode zal echter traag zijn en waarschijnlijk niet robuust.

elementen gerapporteerd worden die in minstens één procent van de spectra voorspeld worden met een zelfzekerheid boven de drempelwaarde bepaald in het vorige hoofdstuk.

4.1 $\text{LaMnO}_3/\text{BaTiO}_3/\text{SrTiO}_3$ superrooster

Figuren 4.1(A) en 4.1(B) tonen een visuele voorstelling van een spectraalbeeld van een $\text{LaMnO}_3/\text{BaTiO}_3/\text{SrTiO}_3$ superrooster dat uitgebreid besproken wordt door Chen et al. [53]. Figuur 4.1(A) toont de gesommeerde intensiteit van elk spectrum. Hierop worden 4 probe-posities aangeduid waarvoor de overeenkomstige spectra getoond worden in figuur 4.1(B). Punt één bevindt zich in het SrTiO_3 gebied, maar de Sr edges bevinden zich niet binnen het opgemeten energieverliesgebied. Punt twee bevindt zich in het BaTiO_3 gebied, punt drie bevindt zich in het overgangsgebied tussen BaTiO_3 en LaMnO_3 en punt vier bevindt zich in het LaMnO_3 gebied. Elementidentificatie wordt gedaan door het $2 \times \text{ViT} + 3 \times \text{U-Net}$ ensemble. Een voorspelling op alle 7826 spectra in dit beeld duurt ongeveer één minuut op een standaard desktop pc met de eerder vermelde GPU. De voorspelde globale elementaire inhoud, volgens het vooropgestelde selectiecriterium, is inderdaad O, Ti, Mn, Ba en La. Figuur 4.1(C) toont, voor alle gevonden elementen en bij elke probe-positie, de voorspelling van het NN en de resultaten van een model-gebaseerde kwantificatie als referentie. Voor het NN wordt de waarschijnlijkheid van aanwezigheid weergegeven. Voor model-gebaseerde kwantificatie wordt de geschatte concentratie van het element weergegeven.

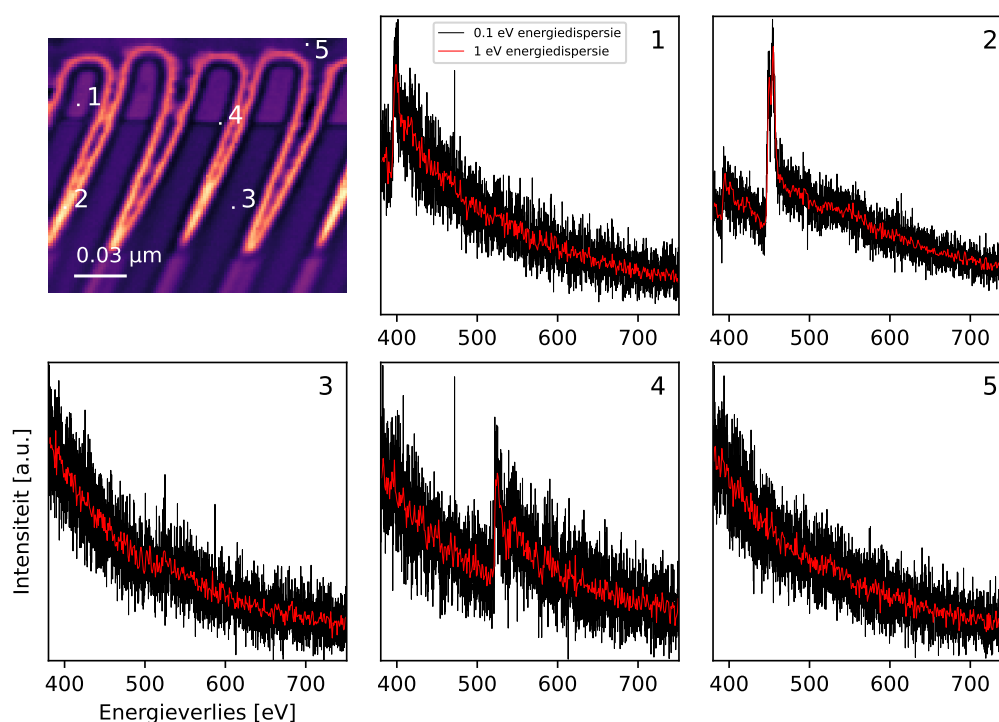


Figuur 4.1. (A) Gesommeerde intensiteit van elk spectrum (B) Spectra bij aangeduide probe-posities (C) Resultaten van elementidentificatie. NN toont per element de waarschijnlijkheid van aanwezigheid, voorspeld door het $2 \times \text{ViT} + 3 \times \text{U-Net}$ ensemble. Ref. toont de resultaten van een model-gebaseerde kwantificatie.

Voor alle elementen bestaat er een onmiskenbare overeenkomst tussen de voorspelling van het NN, model-gebaseerde kwantificatie en visuele inspectie. Spectrum twee toont visueel geen duidelijke aanwezigheid van La, en ook het NN voorspelt geen aanwezigheid van La in spectrum twee. Model-gebaseerde kwantificatie meet wel een zekere La aanwezigheid in het BaTiO_3 gebied, omdat het de overlappende edges van Ba en La niet perfect kan onderscheiden. Er kan besloten worden dat het NN inderdaad een $\text{LaMnO}_3/\text{BaTiO}_3/\text{TiO}_3$ superrooster-structuur voorspelt die goed aansluit bij de resultaten van een visuele inspectie en van model-gebaseerde kwantificatie.

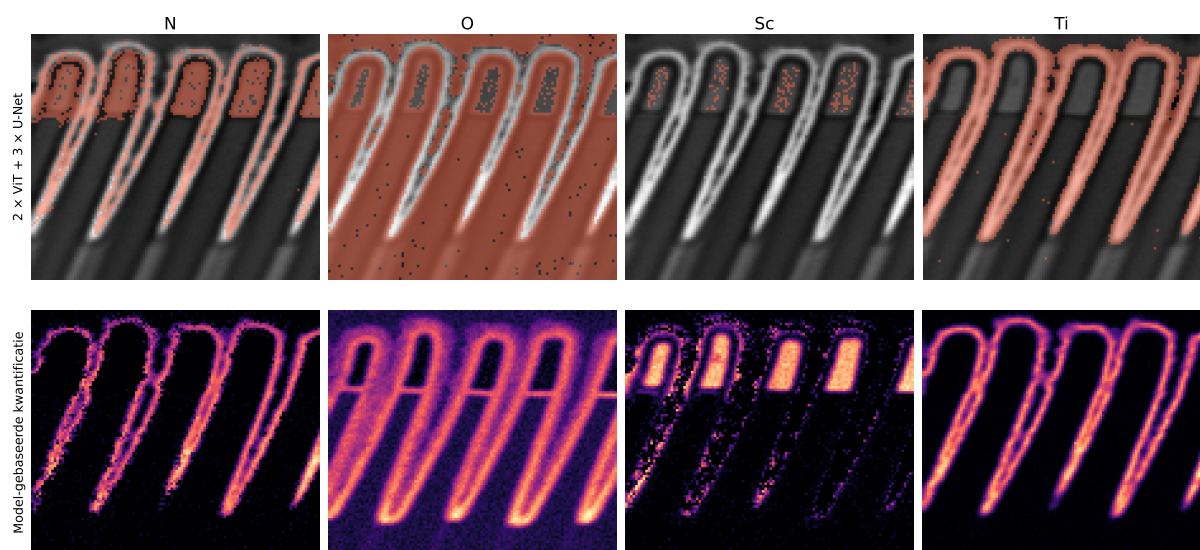
4.2 Ti-O-N sample

Het spectraalbeeld van het Ti-O-N sample wordt samengevat voorgesteld in figuur 4.2. De kleurenmap links bovenaan toont de gesommeerde intensiteit van elk spectrum. Hierop worden vijf probe-posities aangeduid waarvoor de overeenkomstige spectra getoond worden. De onbewerkte spectra, afgebeeld in het zwart, hebben een energieresolutie van 0.1 eV. De vereiste energieresolutie is 1 eV. Hiertoe wordt de energie-as opnieuw gebinned en een kubische spline gebruikt om elk spectrum te evalueren bij de gewenste energie-as. Dit resulteert in de bewerkte spectra afgebeeld in het rood.



Figuur 4.2. Voorstelling van Ti-O-N sample. Het spectrum met 1 eV energieresolutie volgt uit het originele spectrum na herbinning en uitlijning.

Figuur 4.3 toont de voorspellingen gemaakt door het $2 \times \text{ViT} + 3 \times \text{U-Net}$ ensemble en de resultaten van model-gebaseerde kwantificatie, met als input de door het NN gevonden elementen, namelijk N, O, Ti en Sc. Voor Ti en O vormt het NN voorspellingen die goed overeenkomen met wat verwacht kan worden uit visuele inspectie en model-gebaseerde kwantificatie. Voor N bestaat er in bepaalde gebieden een duidelijke verwarring met Sc. Dit is enigszins te verwachten daar dit volgens de eerdere verwarringsmatrix de vaakst verwarde combinatie tussen alle 80 betrokken elementen is. Het NN heeft echter wel een duidelijke voorkeur voor het correcte element en voorspelt slechts in beperkte gebieden een vals positie aanwezigheid, naast de echt positieve voorspelling. Bovendien kan ook model-gebaseerde kwantificatie N en Sc niet correct onderscheiden. Een expert zal – in dit voorbeeld en bij gebrek aan verdere context – hoogstwaarschijnlijk ook moeilijkheden ondervinden in het onderscheiden van N en Sc. Deze ambiguïteit over edges die voorkomen bij dezelfde onset energie is inherent aan EELS eerder dan een tekortkoming van het NN.



Figuur 4.3. Voorspellingen van het $2 \times \text{ViT} + 3 \times \text{U-Net}$ ensemble en model-gebaseerde kwantificatie op het spectraal beeld van het Ti-O-N sample.

Bespreking

Bij aanvang van de dataset-simulatie was het onzeker in hoeverre gesimuleerde EELS spectra op basis van generieke fijnstructuren generalisatie naar experimentele spectra zouden toelaten. Uit de resultaten besproken in de twee voorgaande hoofdstukken kan besloten worden dat de gesimuleerde spectra wel degelijk een doeltreffende generalisatie toelaten. Naast de mogelijkheden in volledig gebruikers-onafhankelijke verwerking van spectra, kunnen de ontwikkelde NN ook nuttig zijn als bv. vertrekpunt of tweede opinie bij een identificatie door een expert. Een grootschalige volledig automatische toepassingen lijkt niet meteen te verwachten, voornamelijk t.g.v. de beperkingen die reeds komen met de opstelling van EELS. De voorspellingen zijn voldoende snel en robuust om een experiment live te ondersteunen. De gepresenteerde NN kunnen nuttig alleenstaand gebruikt worden, maar vormen verder een krachtig hulpmiddel in verdere automatisering van model-gebaseerde kwantificatie. Model-gebaseerde kwantificatie heeft nog steeds het voordeel van een hogere flexibiliteit en meer verwerkingsmogelijkheden, maar vereist in ruil hiervoor een zekere manuele optimalisatie door de gebruiker. De deep learning voorspellingen hebben een beperkte flexibiliteit, maar vereisen geen enkele input van de gebruiker.

Zoals met alle vormen van artificiële intelligentie, is het aangeraden om toch te reflecteren op de voorspellingen wanneer deze gebruikt worden in toepassingen. Eén aandachtspunt waar tussenkomst van een expert wenselijk blijft, is bij het gebruik van een drempelwaarde. In veel gevallen zijn de aanwezige elementen inderdaad die met de hoogste voorspelde waarschijnlijkheid, maar is de output na drempelwaardering strikt gezien niet correct. Het NN zelf heeft echter geen kennis van een drempelwaarde die gebruikt wordt om een voorspelling als aanwezig of afwezig te beschouwen. Een goed voorbeeld hiervan is opnieuw de verwarring van N voor Sc zoals in het Ti-O-N spectraalbeeld. In de meerderheid van de gevallen zal het NN correct N als meest waarschijnlijke element voorspellen. Daar de waarschijnlijkheid dat de geobserveerde edge eerder door Sc veroorzaakt werd ook significant is, kan dit element ook een hoge waarschijnlijkheid toegekend krijgen. Indien de waarschijnlijkheid die aan Sc toegekend wordt hoger is dan de gebruikte drempelwaarde, volgt strikt gezien een verkeerde voorspellingen. Het NN kan niet reflecteren over het idee dat hierbij één geobserveerde edge wordt geïnterpreteerd als veroorzaakt door zowel N en Sc, wat onwaarschijnlijk is. Een expert kan dit wel, en zal dankzij een beter gesofisticeerde interpretatie van de onbewerkte output van het NN wel het correcte resultaat bekomen. In de voorgaande hoofdstukken

wordt veel aandacht besteed aan de voorspellingen van het NN na drempelwaardering omdat deze toelaten de resultaten te kwantificeren. Toch moet benadrukt worden dat de onbewerkte output van het NN de meest waardevolle vorm van informatie is. Men mag niet vergeten dat in toepassingen waar toch een expert betrokken is, het NN hier nog steeds een waardevol hulpmiddel vormt.

Conclusie en verder onderzoek

In deze thesis wordt getracht de mogelijkheden in volledig automatische verwerking van elektronen energieverlies spectra te verbeteren door middel van deep learning technieken. In het bijzonder wordt er getracht een neurale netwerk te ontwikkelen dat de elementaire inhoud van een core-loss spectrum kan voorspellen. Hiertoe werd een dataset met 736 000 gelabelde EELS spectra gesimuleerd op basis van publiek beschikbare generalised oscillator strength tabellen. Het gebruik van volledig gesimuleerde spectra laat toe om een zeer gevarieerde dataset te simuleren die een breed gamma aan elementen en edges representeert. De dataset hier voorgesteld representeert 107 K, L, M of N edges in de 80 elementen van Be tot en met Bi. Generieke fijnstructuren en low-loss gebieden worden gebruikt om o.a. vastestof-effecten en meervoudige verstrooiing na te bootsen, zonder gebruik van kostelijke sample-afhankelijke berekeningen.

De gesimuleerde dataset wordt gebruikt voor het trainen en evalueren van zes verschillende neurale netwerk architecturen, namelijk een meerlaags perceptron, een convolutioneel neurale netwerk, een U-Net, een residueel neurale netwerk, een vision transformer en een compacte convolutionele transformer. Deze netwerken werden vergeleken in termen van identificatievermogen en efficiëntie op zowel gesimuleerde testdata als op experimentele testdata verzameld uit online databanken. Het U-Net en de vision transformer tonen de beste generalisatie naar experimentele data. Een ensemble bestaande uit drie U-Net's en twee vision transformers behaalt een F_1 -score van 0.86, een precisie van 0.88, een sensitiviteit van 0.88 en een exact match ratio van 0.70 op een experimentele testset.

De toepassing van deze neurale netwerken in volledig geautomatiseerde verwerking van twee EELS spectraalbeelden werd gedemonstreerd. Het neurale netwerk produceert snel en betrouwbaar annotaties van de aanwezigheid van elementen in vergelijking met menselijke en model-gebaseerde annotaties.

Daar de technieken in deep learning aan een sneltempo vernieuwen en verbeteren, is het zeer waarschijnlijk dat de neurale netwerken hier voorgesteld al snel vervangbaar zullen zijn door efficiëntere en correctere alternatieven. In dat opzicht vormt de beschikbaarheid van de gesimuleerde dataset – die generalisatie naar experimentele data toelaat – hier een essentiële bijdrage. De gesimuleerde dataset bevat ook informatie over de verhoudingen waarin elementen voorkomen. Dit zou in de toekomst gebruikt

kunnen worden om een neuraal netwerk tijdens de training meer informatie te geven, wat deze nog performanter zou kunnen maken. Verder kan nog veel vooruitgang geboekt worden door de verwerking van de output van het neurale netwerk verder te sofisticeren. Drempelwaardeing heeft zich nuttig bewezen, maar maakt bv. geen gebruik van gekende veelvoorkomende verwarringen. Nieuwe technieken, die deze bijkomende informatie verwerken in de interpretatie van de output van het neurale netwerk, vormen een interessant pad in de verdere automatisering van EELS identificatie.

Bibliografie

- [1] Gatan. “EELS Atlas”. (2023), adres: <https://eels.info/atlas> (bezocht op 13-05-2023).
- [2] J. Verbeeck en S. Van Aert, “Model based quantification of EELS spectra”, *Ultramicroscopy*, jrg. 101, nr. 2, p. 207–224, 2004, ISSN: 0304-3991. DOI: [10.1016/j.ultramic.2004.06.004](https://doi.org/10.1016/j.ultramic.2004.06.004).
- [3] I. Goodfellow, Y. Bengio en A. Courville, *Deep Learning*. MIT Press, 2016, <http://www.deeplearningbook.org>.
- [4] C. C. Aggarwal, *Neural Networks and Deep Learning, A Textbook*, 1ste ed. Springer, 2018, ISBN: 978-3-319-94462-3. DOI: [10.1007/978-3-319-94463-0](https://doi.org/10.1007/978-3-319-94463-0).
- [5] M. Kundmann en O. Krivanek, “Automated processing of parallel-detection EELS data”, *Microsc. Microanal. Microstruct.*, jrg. 2, p. 245–256, 1991. DOI: [10.1051/mmm:0199100202-3024500](https://doi.org/10.1051/mmm:0199100202-3024500).
- [6] G. Kothleitner en F. Hofer, “Elemental occurrence maps: a starting point for quantitative EELS spectrum image processing”, *Ultramicroscopy*, jrg. 96, nr. 3, p. 491–508, 2003, Proceedings of the International Workshop on Strategies and Advances in Atomic Level Spectroscopy and Analysis, ISSN: 0304-3991. DOI: [10.1016/S0304-3991\(03\)00111-6](https://doi.org/10.1016/S0304-3991(03)00111-6).
- [7] M. Shiga, S. Muto, K. Tatsumi en K. Tsuda, “Matrix Factorization for Automatic Chemical Mapping from Electron Microscopic Spectral Imaging Datasets”, *Transactions of the Materials Research Society of Japan*, jrg. 41, nr. 4, p. 333–336, 2016. DOI: [10.14723/tmrsj.41.333](https://doi.org/10.14723/tmrsj.41.333).
- [8] M.-T. Chang, R.-F. Cai, C.-C. Chen en S.-C. Lo, “Development of Clustering Algorithm Applied for the EELS Analysis of Advanced Devices”, *Microscopy and Microanalysis*, jrg. 26, nr. S2, p. 2112–2114, 2020. DOI: [10.1017/S1431927620020486](https://doi.org/10.1017/S1431927620020486).

- [9] J. Hachtel, N. Borodinov, K. Roccapriore e.a., “Beyond NMF: Advanced Signal Processing and Machine Learning Methodologies for Hyperspectral Analysis in EELS”, *Microscopy and Microanalysis*, jrg. 27, nr. S1, p. 322–324, 2021. DOI: [10.1017/S1431927621001720](https://doi.org/10.1017/S1431927621001720).
- [10] T. Blum, J. Graves, M. Zachman, R. Kannan, X. Pan en M. Chi, “Machine Learning for Challenging EELS and EDS Spectral Decomposition”, *Microscopy and Microanalysis*, jrg. 25, nr. S2, p. 180–181, 2019. DOI: [10.1017/S1431927619001636](https://doi.org/10.1017/S1431927619001636).
- [11] M. Chatzidakis en G. Botton, “Towards calibration-invariant spectroscopy using deep learning”, *Scientific Reports*, jrg. 9, p. 2126, feb 2019. DOI: [10.1038/s41598-019-38482-1](https://doi.org/10.1038/s41598-019-38482-1).
- [12] D. del-Pozo-Bueno, F. Peiró en S. Estradé, “Support vector machine for EELS oxidation state determination”, *Ultramicroscopy*, jrg. 221, p. 113 190, 2021, ISSN: 0304-3991. DOI: [10.1016/j.ultramicro.2020.113190](https://doi.org/10.1016/j.ultramicro.2020.113190).
- [13] C. Pate, J. Hart en M. Taheri, “RapidEELS: machine learning for denoising and classification in rapid acquisition electron energy loss spectroscopy”, *Scientific Reports*, jrg. 11, sep 2021. DOI: [10.1038/s41598-021-97668-8](https://doi.org/10.1038/s41598-021-97668-8).
- [14] L. I. Roest, S. E. van Heijst, L. Maduro, J. Rojo en S. Conesa-Boj, “Charting the low-loss region in electron energy loss spectroscopy with machine learning”, *Ultramicroscopy*, jrg. 222, p. 113 202, 2021, ISSN: 0304-3991. DOI: [10.1016/j.ultramicro.2021.113202](https://doi.org/10.1016/j.ultramicro.2021.113202).
- [15] S. Shayan Mousavi M, A. Pofelski en G. Botton, “EELSpecNet: Deep Convolutional Neural Network Solution for Electron Energy Loss Spectroscopy Deconvolution”, *Microscopy and Microanalysis*, jrg. 27, nr. S1, p. 1626–1627, 2021. DOI: [10.1017/S1431927621005997](https://doi.org/10.1017/S1431927621005997).
- [16] N. R. Lewis, Y. Jin, X. Tang e.a., “Forecasting of in situ electron energy loss spectroscopy”, *npj Computational Materials*, jrg. 8, 1 dec 2022, ISSN: 20573960. DOI: [10.1038/s41524-022-00940-2](https://doi.org/10.1038/s41524-022-00940-2).
- [17] Y. Sun, S. Brockhauser en P. Hegedűs, “Comparing End-to-End Machine Learning Methods for Spectra Classification”, *Applied Sciences*, jrg. 11, nr. 23, 2021, ISSN: 2076-3417. DOI: [10.3390/app112311520](https://doi.org/10.3390/app112311520).
- [18] J. Acquarelli, T. van Laarhoven, J. Gerretzen, T. N. Tran, L. M. Buydens en E. Marchiori, “Convolutional neural networks for vibrational spectroscopic data

- analysis”, *Analytica Chimica Acta*, jrg. 954, p. 22–31, 2017, ISSN: 0003-2670. DOI: [10.1016/j.aca.2016.12.010](https://doi.org/10.1016/j.aca.2016.12.010).
- [19] M. Kazemzadeh, C. L. Hisey, K. Zargar-Shoshtari, W. Xu en N. G. Broderick, “Deep convolutional neural networks as a unified solution for Raman spectroscopy-based classification in biomedical applications”, *Optics Communications*, jrg. 510, p. 127 977, 2022, ISSN: 0030-4018. DOI: [10.1016/j.optcom.2022.127977](https://doi.org/10.1016/j.optcom.2022.127977).
- [20] L.-N. Li, X.-F. Liu, F. Yang, W.-M. Xu, J.-Y. Wang en R. Shu, “A review of artificial neural network based chemometrics applied in laser-induced breakdown spectroscopy analysis”, *Spectrochimica Acta Part B: Atomic Spectroscopy*, jrg. 180, p. 106 183, 2021, ISSN: 0584-8547. DOI: [10.1016/j.sab.2021.106183](https://doi.org/10.1016/j.sab.2021.106183).
- [21] C. Jones, N. Daly, C. Higgitt en M. Rodrigues, “Neural network-based classification of X-ray fluorescence spectra of artists’ pigments: an approach leveraging a synthetic dataset created using the fundamental parameters method”, *Heritage Science*, jrg. 10, jun 2022. DOI: [10.1186/s40494-022-00716-3](https://doi.org/10.1186/s40494-022-00716-3).
- [22] C. Li, D. Wang en L. Kong, “Application of Machine Learning Techniques in Mineral Classification for Scanning Electron Microscopy - Energy Dispersive X-Ray Spectroscopy (SEM-EDS) Images”, *Journal of Petroleum Science and Engineering*, jrg. 200, p. 108 178, 2021, ISSN: 0920-4105. DOI: [10.1016/j.petrol.2020.108178](https://doi.org/10.1016/j.petrol.2020.108178).
- [23] K. Ghosh, A. Stuke, M. Todorović e.a., “Deep Learning Spectroscopy: Neural Networks for Molecular Excitation Spectra”, *Advanced Science*, jrg. 6, nr. 9, p. 1 801 367, 2019. DOI: [10.1002/advs.201801367](https://doi.org/10.1002/advs.201801367).
- [24] L. Kong, Z. Ji en H. Xin, “Electron energy loss spectroscopy database synthesis and automation of core-loss edge recognition by deep-learning neural networks”, *Scientific Reports*, jrg. 12, p. 22 183, 2022. DOI: [10.1038/s41598-022-25870-3](https://doi.org/10.1038/s41598-022-25870-3).
- [25] R. Egerton, *Electron Energy-Loss Spectroscopy in the Electron Microscope, 3rd Edition* (Springer, 2011). jan 2011, ISBN: 978-1-4419-9582-7. DOI: [10.1007/978-1-4419-9583-4](https://doi.org/10.1007/978-1-4419-9583-4).
- [26] M. P. Seah en W. A. Dench, “Quantitative electron spectroscopy of surfaces: A standard data base for electron inelastic mean free paths in solids”, *Surface and Interface Analysis*, jrg. 1, p. 2–11, 1979.

- [27] A. Jain, S. P. Ong, G. Hautier e.a., “Commentary: The Materials Project: A materials genome approach to accelerating materials innovation”, *APL Materials*, jrg. 1, nr. 1, p. 011 002, 2013. DOI: [10.1063/1.4812323](https://doi.org/10.1063/1.4812323).
- [28] Z. Zhang, I. Lobato, D. Jannis, J. Verbeeck, S. V. Aert en P. Nellist, *Generalised oscillator strength for core-shell electron excitation by fast electrons based on Dirac solutions*, versie 1.0, Zenodo, mrt 2023. DOI: [10.5281/zenodo.7729585](https://doi.org/10.5281/zenodo.7729585). adres: <https://doi.org/10.5281/zenodo.7729585>.
- [29] J. Verbeeck en D. Jannis, *pyEELSmodel (onuitgebracht)*, 2023.
- [30] J. Verbeeck, S. Van Aert en G. Bertoni, “Model-based quantification of EELS spectra: Including the fine structure”, *Ultramicroscopy*, jrg. 106, nr. 11, p. 976–980, 2006, Proceedings of the International Workshop on Enhanced Data Generated by Electrons, ISSN: 0304-3991. DOI: [10.1016/j.ultramicro.2006.05.006](https://doi.org/10.1016/j.ultramicro.2006.05.006).
- [31] J. Chae, J.-S. Kim, S.-Y. Nam, M. Kim en J. Park, “Introduction to the standard reference data of electron energy loss spectra and their database: eel.geri.re.kr”, *Applied Microscopy*, jrg. 50, dec 2020. DOI: [10.1186/s42649-019-0015-3](https://doi.org/10.1186/s42649-019-0015-3).
- [32] P. Ewels, T. Sikora, V. Serin, C. P. Ewels en L. Lajaunie, “A Complete Overhaul of the Electron Energy-Loss Spectroscopy and X-Ray Absorption Spectroscopy Database: eelsdb.eu”, *Microscopy and Microanalysis*, jrg. 22, p. 717–724, feb 2016, ISSN: 1435-8115. DOI: [10.1017/S1431927616000179](https://doi.org/10.1017/S1431927616000179).
- [33] W. Tirry, D. Schryvers, K. Jorissen en D. Lamoen, “Electron-diffraction structure refinement of Ni₄Ti₃ precipitates in Ni₅₂Ti₄₈”, *Acta Crystallographica Section B-structural Science - ACTA CRYSTALLOGR B-STRUCT SCI*, jrg. 62, p. 966–971, dec 2006. DOI: [10.1107/S0108768106036457/lc5049sup2.txt](https://doi.org/10.1107/S0108768106036457/lc5049sup2.txt).
- [34] A. Feldhoff, J. Martynczuk, M. Arnold e.a., “Spin-state transition of iron in (Ba_{0.5}Sr_{0.5})(Fe_{0.8}Zn_{0.2})O₃- perovskite”, *Journal of Solid State Chemistry*, jrg. 182, nr. 11, p. 2961–2971, 2009, ISSN: 0022-4596. DOI: [10.1016/j.jssc.2009.07.058](https://doi.org/10.1016/j.jssc.2009.07.058).
- [35] L. Lajaunie, F. Boucher, R. Dessapt en P. Moreau, “Quantitative use of electron energy-loss spectroscopy Mo-M_{2,3} edges for the study of molybdenum oxides”, *Ultramicroscopy*, jrg. 149, p. 1–8, 2015, ISSN: 0304-3991. DOI: [10.1016/j.ultramicro.2014.11.002](https://doi.org/10.1016/j.ultramicro.2014.11.002).

- [36] F. Hofer en P. Golob, “New examples for near-edge fine structures in electron energy loss spectroscopy”, *Ultramicroscopy*, jrg. 21, nr. 4, p. 379–383, 1987, ISSN: 0304-3991. DOI: [10.1016/0304-3991\(87\)90036-2](https://doi.org/10.1016/0304-3991(87)90036-2).
- [37] G. Gassner, J. Patscheider, P. H. Mayrhofer e.a., “Structure of sputtered nanocomposite CrC_x-C:H thin films”, *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures Processing, Measurement, and Phenomena*, jrg. 24, nr. 4, p. 1837–1843, 2006. DOI: [10.1116/1.2216713](https://doi.org/10.1116/1.2216713).
- [38] G. Hug, R. Leapman en A. Jaouen, “A study of EELS fine structure in three chromium carbide”, *Microbeam Analysis*, 1995.
- [39] S. Trasobares, C. Kolczewski, R. Rätty e.a., “Monitoring the Decomposition of Melamine in the Solid Phase by Electron Energy Loss Chronospectroscopy”, *The Journal of Physical Chemistry A*, jrg. 107, nr. 2, p. 228–235, 2003. DOI: [10.1021/jp026340j](https://doi.org/10.1021/jp026340j).
- [40] V. Serin, C. Colliex, R. Brydson, S. Matar en F. Boucher, “EELS investigations of the electron conduction band states in wurtzite AlN and oxygen-doped AlN(O)”, *Physical Review*, 1998.
- [41] A. Krizhevsky, I. Sutskever en G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks”, in *Advances in Neural Information Processing Systems*, F. Pereira, C. Burges, L. Bottou en K. Weinberger, red., deel 25, Curran Associates, Inc., 2012. adres: <https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf>.
- [42] K. Simonyan en A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition”, 3rd International Conference on Learning Representations (ICLR 2015), 2015, p. 1–14.
- [43] O. Ronneberger, P. Fischer en T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation”, in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells en A. F. Frangi, red., Cham: Springer International Publishing, 2015, p. 234–241, ISBN: 978-3-319-24574-4.
- [44] K. He, X. Zhang, S. Ren en J. Sun, “Deep Residual Learning for Image Recognition”, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Los Alamitos, CA, USA: IEEE Computer Society, jun 2016, p. 770–778. DOI: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90).

- [45] H. Iqbal, *HarisIqbal88/PlotNeuralNet v1.0.0*, versie v1.0.0, dec 2018. DOI: [10.5281/zenodo.2526396](https://doi.org/10.5281/zenodo.2526396). adres: <https://doi.org/10.5281/zenodo.2526396>.
- [46] S. Hochreiter en J. Schmidhuber, “Long Short-term Memory”, *Neural computation*, jrg. 9, p. 1735–80, dec 1997. DOI: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735).
- [47] A. Vaswani, N. Shazeer, N. Parmar e.a., “Attention is All you Need”, in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio e.a., red., deel 30, Curran Associates, Inc., 2017. adres: <https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf>.
- [48] T. Brown, B. Mann, N. Ryder e.a., “Language Models are Few-Shot Learners”, in *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan en H. Lin, red., deel 33, Curran Associates, Inc., 2020, p. 1877–1901. adres: <https://proceedings.neurips.cc/paper/2020/file/1457c0d6bfc4967418bfb8ac142f64a-Paper.pdf>.
- [49] A. Dosovitskiy, L. Beyer, A. Kolesnikov e.a., “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale”, 9th International Conference on Learning Representations (ICLR 2021), 2021.
- [50] A. Hassani, S. Walton, N. Shah, A. Abuduweili, J. Li en H. Shi, *Escaping the Big Data Paradigm with Compact Transformers*, 2021. DOI: [10.48550/ARXIV.2104.05704](https://doi.org/10.48550/ARXIV.2104.05704).
- [51] R. Battiti en A. M. Colla, “Democracy in neural nets: Voting schemes for classification”, *Neural Networks*, jrg. 7, nr. 4, p. 691–707, 1994, ISSN: 0893-6080. DOI: [https://doi.org/10.1016/0893-6080\(94\)90046-9](https://doi.org/10.1016/0893-6080(94)90046-9).
- [52] R. R. Selvaraju, A. Das, R. Vedantam, M. Cogswell, D. Parikh en D. Batra, “Grad-CAM: Why did you say that? Visual Explanations from Deep Networks via Gradient-based Localization”, *CoRR*, jrg. abs/1610.02391, 2016. DOI: [1610.02391](https://doi.org/10.48550/ARXIV.1610.02391). adres: <http://arxiv.org/abs/1610.02391>.
- [53] B. Chen, N. Gauquelin, N. Strkalj e.a., “Signatures of enhanced out-of-plane polarization in asymmetric BaTiO₃ superlattices integrated on silicon”, *Nature Communications*, jrg. 13, jan 2022. DOI: [10.1038/s41467-021-27898-x](https://doi.org/10.1038/s41467-021-27898-x).

Data en code beschikbaarheid

De gesimuleerde dataset hier voorgesteld zal in de nabije toekomst openbaar beschikbaar worden gemaakt op op Zenodo. De code omtrent het simuleren van de dataset wordt openbaar beschikbaar gemaakt wanneer de onderliggende software pyEELSmodel openbaar beschikbaar wordt. De getrainde neurale netwerken en software omtrent het maken en gebruiken van deze wordt in de nabije toekomst beschikbaar gemaakt op GitHub.

Dankwoord

Ik wil graag mijn promotor Prof. Johan Verbeeck bedanken voor de fijne samenwerking de voorbije jaren. Het is bijzonder fijn samenwerken met iemand die keer op keer boeiende ideeën weet aan te brengen. Vervolgens wil ik graag mijn co-promotor Dr. Daen Jannis bedanken. Ik kon de voorbije jaren steeds bij hem terecht voor hulp, ideeën en feedback. Daarnaast wil ik ook graag Dr. Ivan Lobato bedanken om mij op weg te helpen in de eerste fase van deze thesis.

Er werd infrastructuur en dienstverlening gebruikt in dit werk, voorzien door het VSC (Vlaams Supercomputer Centrum), gefinancierd door het FWO en de Vlaamse overheid. Verder wil ik graag Dr. Bert Tijskens bedanken voor de hulp bij het optimaliseren van de HPC-code.

Appendix

A.1 Benoemingconventies van core-loss edges

Tabel A.1. Benoemingconventies van core-loss edges, afkomstig van Egerton [25]

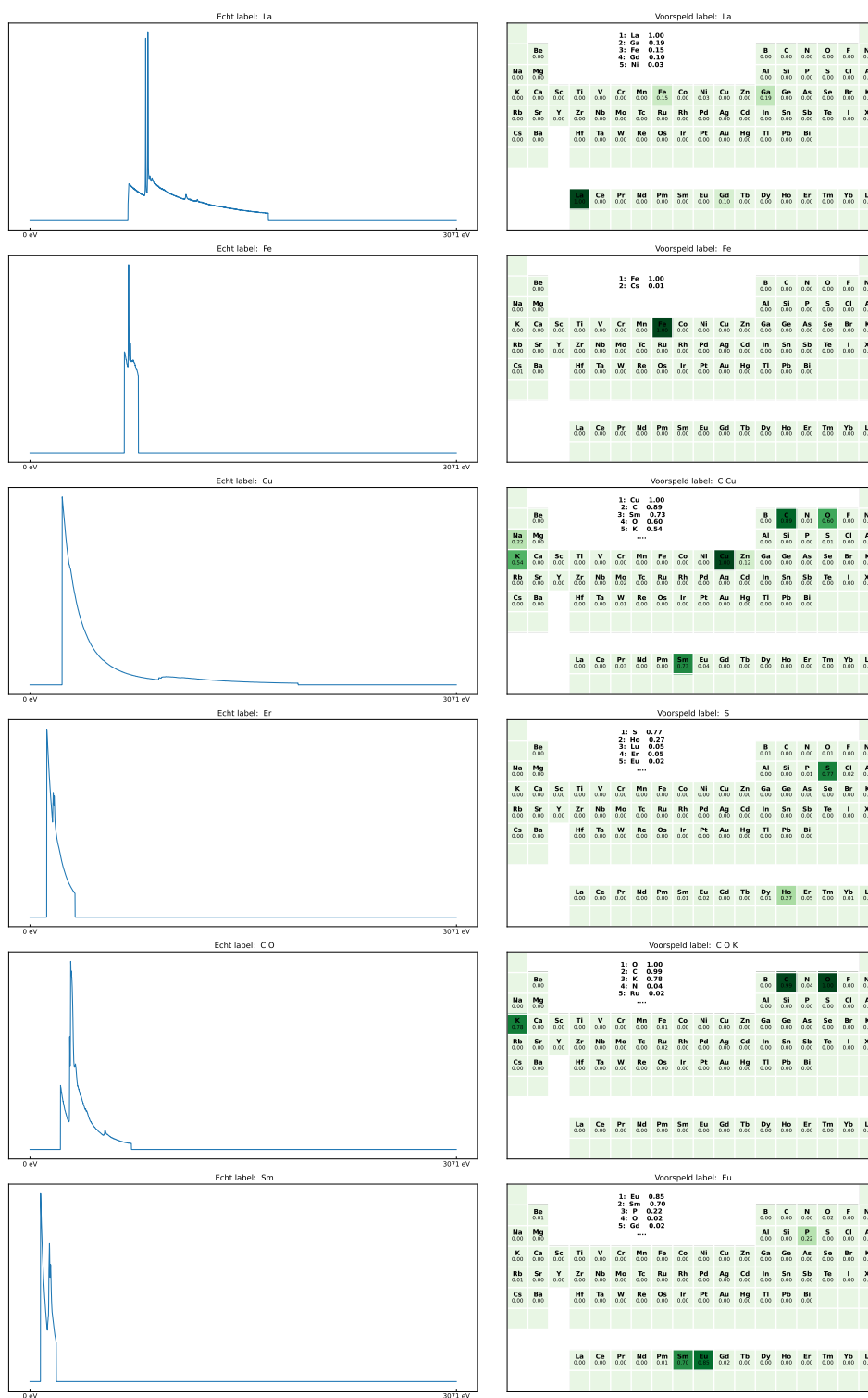
Naam	Toestand	n	l	j
K	$1s^{1/2}$	1	0	1/2
L ₁	$2s^{1/2}$	2	0	1/2
L ₂	$2p^{1/2}$	2	1	1/2
L ₃	$2p^{3/2}$	2	1	3/2
M ₁	$3s^{1/2}$	3	0	1/2
M ₂	$3p^{1/2}$	3	1	1/2
M ₃	$3p^{3/2}$	3	1	3/2
M ₄	$3d^{3/2}$	3	2	3/2
M ₅	$3d^{5/2}$	3	2	5/2
N ₁	$4s^{1/2}$	4	0	1/2
N ₂	$4p^{1/2}$	4	1	1/2
N ₃	$4p^{3/2}$	4	1	3/2
N ₄	$4d^{3/2}$	4	2	3/2
N ₅	$4d^{5/2}$	4	2	5/2
N ₆	$4f^{5/2}$	4	3	5/2
N ₇	$4f^{7/2}$	4	3	7/2
O ₁	$5s^{1/2}$	5	0	1/2
O ₂	$5p^{1/2}$	5	1	1/2
O ₃	$5p^{3/2}$	5	1	3/2
O ₄	$5d^{3/2}$	5	2	3/2
O ₅	$5p^{5/2}$	5	2	5/2

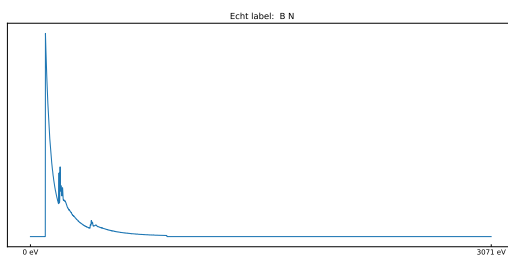
A.2 Evaluatie van extra U-Net, ViT en CTT instanties.

Tabel A.2. Vergelijking van identificatie architecturen bij gebruik van drempelwaarde bepaald i.f.v. experimentele spectra

Model	Instantie	Exp. test F₁-score	Exp. test EMR	Drempel- waarde
U-Net	1	0.81	0.63	0.90
	2	0.84	0.61	0.90
	3	0.82	0.63	0.95
ViT	1	0.83	0.62	0.85
	2	0.84	0.60	0.80
	3	0.83	0.62	0.80
CCT	1	0.79	0.58	0.85
	2	0.80	0.50	0.85
	3	0.82	0.56	0.85

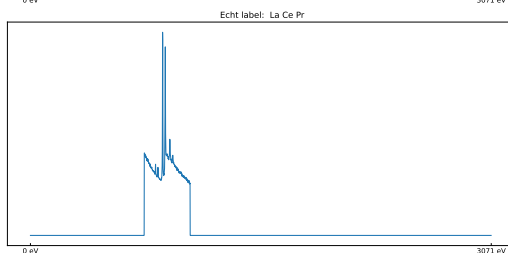
A.3 Voorspellingen op experimentele spectra





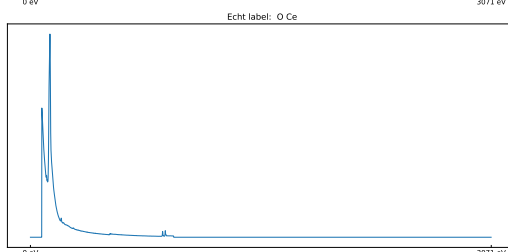
Voorspeld label: B N

1: B 1.00		C 0.04		O 0.00		F 0.00		Ne 0.00	
2: N 0.97		Si 0.34		P 0.00		S 0.00		Cl 0.00	
3: C 0.04		Al 0.00		Ar 0.00		K 0.00		Ca 0.00	
4: Cl 0.04		Sc 0.00		Ti 0.00		V 0.00		Cr 0.00	
.....		Mn 0.00		Fe 0.00		Co 0.00		Ni 0.00	
		Cu 0.00		Zn 0.00		Ga 0.00		Ge 0.00	
		As 0.00		Se 0.00		Br 0.00		Kr 0.00	
		Rb 0.00		Sr 0.00		Y 0.00		Zr 0.00	
		Nb 0.00		Mo 0.00		Tc 0.00		Ru 0.00	
		Rh 0.00		Pd 0.00		Ag 0.00		Cd 0.00	
		In 0.00		Sn 0.00		Sb 0.00		Te 0.00	
		I 0.00		Xe 0.00		Cs 0.00		Ba 0.00	
		Hf 0.00		Ta 0.00		W 0.00		Re 0.00	
		Os 0.00		Ir 0.00		Pt 0.00		Au 0.00	
		Hg 0.00		Tl 0.00		Pb 0.00		Bi 0.00	
		La 0.00		Ce 0.00		Pr 0.00		Nd 0.00	
		Pm 0.00		Sm 0.00		Eu 0.00		Gd 0.00	
		Tb 0.00		Dy 0.00		Ho 0.00		Er 0.00	
		Tm 0.00		Yb 0.00		Lu 0.00			



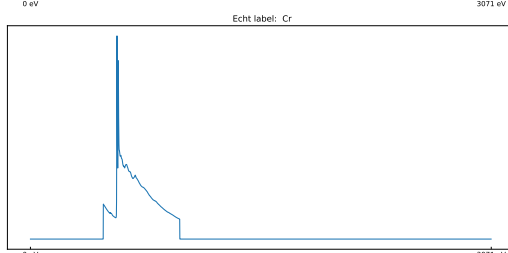
Voorspeld label: Ne

1: Ne 0.98		B 0.00		C 0.00		N 0.00		O 0.00	
2: Ce 0.23		Al 0.00		Si 0.00		P 0.00		S 0.00	
3: La 0.01		Ar 0.00		K 0.00		Ca 0.00		Sc 0.00	
4: Cu 0.01		Ti 0.00		V 0.00		Cr 0.00		Mn 0.00	
		Fe 0.00		Co 0.00		Ni 0.00		Cu 0.00	
		Zn 0.00		Ga 0.00		Ge 0.00		As 0.00	
		Se 0.00		Br 0.00		Kr 0.00		Rb 0.00	
		Sr 0.00		Y 0.00		Zr 0.00		Nb 0.00	
		Mo 0.00		Tc 0.00		Ru 0.00		Rh 0.00	
		Pd 0.00		Ag 0.00		Cd 0.00		In 0.00	
		Sn 0.00		Sb 0.00		Te 0.00		I 0.00	
		Xe 0.00		Cs 0.00		Ba 0.00		Hf 0.00	
		Ta 0.00		W 0.00		Re 0.00		Os 0.00	
		Ir 0.00		Pt 0.00		Au 0.00		Hg 0.00	
		Tl 0.00		Pb 0.00		Bi 0.00		La 0.00	
		Ce 0.00		Pr 0.00		Nd 0.00		Pm 0.00	
		Sm 0.00		Eu 0.00		Gd 0.00		Tb 0.00	
		Dy 0.00		Ho 0.00		Er 0.00		Tm 0.00	
		Yb 0.00		Lu 0.00					



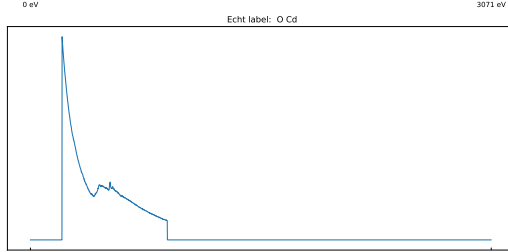
Voorspeld label: O

1: O 0.97		B 0.00		C 0.00		N 0.00		F 0.00	
2: Ce 0.50		Al 0.00		Si 0.00		P 0.00		S 0.00	
3: Fe 0.38		Ar 0.00		K 0.00		Ca 0.00		Sc 0.00	
4: Ni 0.32		Ti 0.00		V 0.00		Cr 0.00		Mn 0.00	
5: Nd 0.15		Fe 0.00		Co 0.00		Ni 0.32		Cu 0.00	
		Zn 0.00		Ga 0.00		Ge 0.00		As 0.00	
		Se 0.00		Br 0.00		Kr 0.00		Rb 0.00	
		Sr 0.00		Y 0.00		Zr 0.00		Nb 0.00	
		Mo 0.00		Tc 0.00		Ru 0.00		Rh 0.00	
		Pd 0.00		Ag 0.00		Cd 0.00		In 0.00	
		Sn 0.00		Sb 0.00		Te 0.00		I 0.00	
		Xe 0.00		Cs 0.00		Ba 0.00		Hf 0.00	
		Ta 0.00		W 0.00		Re 0.00		Os 0.00	
		Ir 0.00		Pt 0.00		Au 0.00		Hg 0.00	
		Tl 0.00		Pb 0.00		Bi 0.00		La 0.00	
		Ce 0.00		Pr 0.00		Nd 0.15		Pm 0.00	
		Sm 0.00		Eu 0.00		Gd 0.00		Tb 0.00	
		Dy 0.00		Ho 0.00		Er 0.00		Tm 0.00	
		Yb 0.00		Lu 0.00					



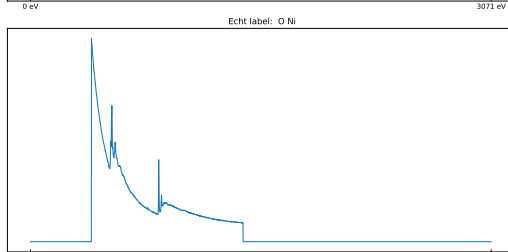
Voorspeld label: Cr

1: Cr 1.00		B 0.00		C 0.00		N 0.00		O 0.00	
2: F 0.56		Al 0.00		Si 0.00		P 0.00		S 0.00	
3: Mn 0.38		Ar 0.00		K 0.00		Ca 0.00		Sc 0.00	
4: O 0.01		Ti 0.00		V 0.00		Cr 0.00		Mn 0.38	
		Fe 0.00		Co 0.00		Ni 0.00		Cu 0.00	
		Zn 0.00		Ga 0.00		Ge 0.00		As 0.00	
		Se 0.00		Br 0.00		Kr 0.00		Rb 0.00	
		Sr 0.00		Y 0.00		Zr 0.00		Nb 0.00	
		Mo 0.00		Tc 0.00		Ru 0.00		Rh 0.00	
		Pd 0.00		Ag 0.00		Cd 0.00		In 0.00	
		Sn 0.00		Sb 0.00		Te 0.00		I 0.00	
		Xe 0.00		Cs 0.00		Ba 0.00		Hf 0.00	
		Ta 0.00		W 0.00		Re 0.00		Os 0.00	
		Ir 0.00		Pt 0.00		Au 0.00		Hg 0.00	
		Tl 0.00		Pb 0.00		Bi 0.00		La 0.00	
		Ce 0.00		Pr 0.00		Nd 0.00		Pm 0.00	
		Sm 0.00		Eu 0.00		Gd 0.00		Tb 0.00	
		Dy 0.00		Ho 0.00		Er 0.00		Tm 0.00	
		Yb 0.00		Lu 0.00					



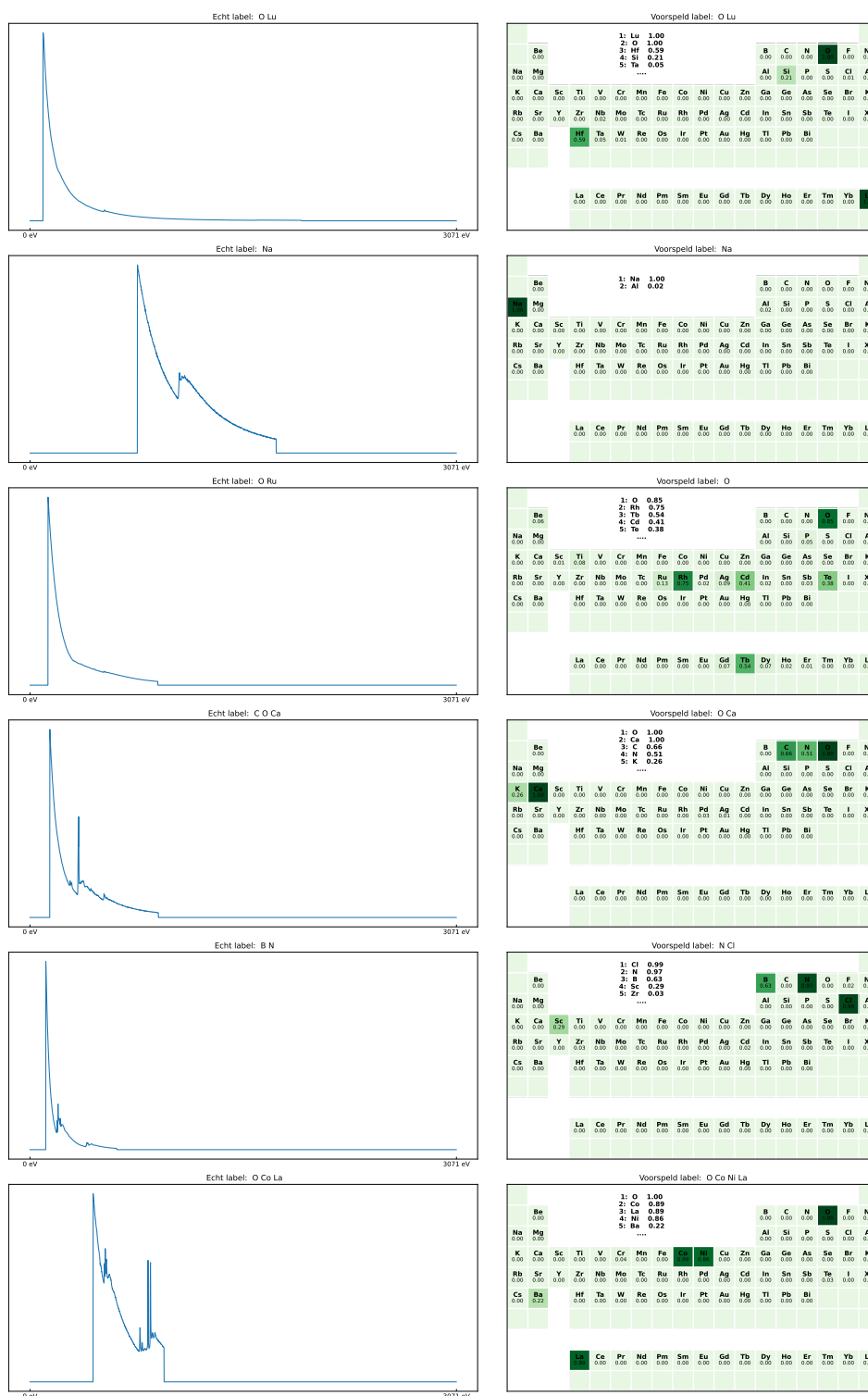
Voorspeld label: O Cd

1: Cd 1.00		B 0.00		C 0.00		N 0.00		F 0.00	
2: O 0.99		Al 0.00		Si 0.00		P 0.00		S 0.00	
3: C 0.29		Ar 0.00		K 0.00		Ca 0.00		Sc 0.00	
4: Sn 0.08		Ti 0.00		V 0.00		Cr 0.00		Mn 0.00	
5: V 0.05		Fe 0.00		Co 0.00		Ni 0.00		Cu 0.00	
		Zn 0.00		Ga 0.00		Ge 0.00		As 0.00	
		Se 0.00		Br 0.00		Kr 0.00		Rb 0.00	
		Sr 0.00		Y 0.00		Zr 0.00		Nb 0.00	
		Mo 0.00		Tc 0.00		Ru 0.00		Rh 0.00	
		Pd 0.00		Ag 0.00		Cd 0.00		In 0.00	
		Sn 0.08		Sb 0.00		Te 0.00		I 0.00	
		Xe 0.00		Cs 0.00		Ba 0.00		Hf 0.00	
		Ta 0.00		W 0.00		Re 0.00		Os 0.00	
		Ir 0.00		Pt 0.00		Au 0.00		Hg 0.00	
		Tl 0.00		Pb 0.00		Bi 0.00		La 0.00	
		Ce 0.00		Pr 0.00		Nd 0.00		Pm 0.00	
		Sm 0.00		Eu 0.00		Gd 0.00		Tb 0.00	
		Dy 0.00		Ho 0.00		Er 0.00		Tm 0.00	
		Yb 0.00		Lu 0.00					



Voorspeld label: O Ni

1: Ni 1.00		B 0.00		C 0.00		N 0.00		F 0.00	
2: O 1.00		Al 0.00		Si 0.00		P 0.00		S 0.00	
3: Mo 0.18		Ar 0.00		K 0.00		Ca 0.00		Sc 0.00	
4: Sb 0.13		Ti 0.00		V 0.00		Cr 0.00		Mn 0.00	
5: Co 0.04		Fe 0.00		Co 0.04		Cu 0.00		Zn 0.00	
		Ga 0.00		Ge 0.00		As 0.00		Se 0.00	
		Br 0.00		Kr 0.00		Rb 0.00		Sr 0.00	
		Y 0.00		Zr 0.00		Nb 0.00		Mo 0.00	
		Tc 0.00		Ru 0.00		Rh 0.00		Pd 0.00	
		Ag 0.00		Cd 0.00		In 0.00		Sn 0.00	
		Sb 0.13		Te 0.00		I 0.00		Xe 0.00	
		Cs 0.00		Ba 0.00		Hf 0.00		Ta 0.00	
		W 0.00		Re 0.00		Os 0.00		Ir 0.00	
		Pt 0.00		Au 0.00		Hg 0.00		Tl 0.00	
		Pb 0.00		Bi 0.00		La 0.00		Ce 0.00	
		Pr 0.00		Nd 0.00		Pm 0.00		Sm 0.00	
		Eu 0.00		Gd 0.00		Tb 0.00		Dy 0.00	
		Ho 0.00		Er 0.00		Tm 0.00		Yb 0.00	
		Lu 0.00							



Figuur A.1. Schematische weergave van de voorspelling van het $2 \times \text{ViT} + 3 \times \text{U-Net}$ ensemble op 30 willekeurige experimentele spectra