Automated Aberration Measurement from Ronchigrams using Machine Learning

James Graca-Jones

Mansfield College

University of Oxford

MEng Materials Science

June 2022

1 Acknowledgments

I thank my supervisors Dr. Chen Huang, for his ongoing support and teaching, and Professor Angus Kirkland, for advising me on the ultimate focus of the project. I also thank Mr. Zhiyuan Ding for verifying the physics of my simulations and Mr. Kevin Treder for assisting me in using the group computer. I thank the University of Oxford for the resources permitting me to complete this project. Finally, I thank my friends and family.

2 Abstract

In STEM, lens aberrations are measured from the Ronchigram by methods failing at low illuminations and that could be exceeded in measurement speed and accuracy. The need to acquire Ronchigrams acceptable for methods like auto-correlation by raising currents or replacing the investigated specimen region with another region of the specimen (or another specimen altogether) with properties conducive to generating the appropriate Ronchigram adds significant time overhead to investigation, hindering progress. Such hindrances pose challenges in achieving real-time aberration measurement and correction; even when the hindrance is decreased by methods that can work on multiple specimen types, high illuminations risk specimen damage by the fine electron probe, limiting what can be discovered from specimens which are transformed by damage. Furthermore, aberration drift between time-consuming measurement setups mean real-time measurement is necessary for maximising STEM resolutions, desirable in all studies. This thesis details attempts to further utilise machine learning, in the form of an *EfficientNet* convolutional neural network, in aberration measurement to combat aforementioned issues. The aim was to train the CNN to, at acceptable speed and accuracy, measure from the Ronchigram magnitudes and angles of depicted aberrations, these predominantly of the probe-forming objective lens. It was aimed to begin with aberrations of 1st and 2nd-order in Krivanek notation and proceed to higher orders. Ultimately, the aim was to enable this method at low illuminations and for varying specimen types via which the Ronchigram is generated, towards real-time aberration measurement. CNNs were finally trained with Ronchigrams simulated to emulate a thin, amorphous specimen in the 300kV C3,0-corrected STEM, in which 3rd-order spherical aberration had magnitudes up to 1um. Evaluation of two-fold astigmatism was prioritised due to its relative instability, this aberration's magnitude and angle predicted by the CNN in test simulations with, for each respective parameter, validation root mean square errors of 15.7% and 18.9% of the highest aberration value shown to the network in training, at an inference time per Ronchigram of 219ms on an Intel® Core™ i9 CPU. Despite requiring further training to extend success to lower illuminations and more measured aberrations, these accuracies may be sufficient where quick, repetitive aberration correction is practicable, and findings of this thesis may inform further methods of aberration measurement using machine learning.

Contents

1 Acknowledgments **2**

2 Abstract **3**

3 Glossary **6**

4 Introduction **8**

4.1 Transmission Electron Microscopy 8

4.2 The Ronchigram and its Aberrations 10

4.3 Machine Learning for Aberration Measurement 14

4.4 Engineering Context 16

5 Literature Review **19**

6 Methodology **23**

6.1 Ronchigram Simulations 23

6.1.1 Simulations Before Detector Noise Application 23

6.1.2 Detector Noise Application to Simulations 23

6.1.3 The Probe 25

6.2 Machine Learning Tools 28

6.3 Network Testing Methods 30

6.4 Incremental Development 31

6.5 Experimental Ronchigram Acquisition 31

7 Discussion of Results **34**

7.1 Training the Network to Evaluate Aberration Constants in Simulations Featuring Only 1st and 2nd Order Aberrations 34

7.1.1 Simulations with a Detector Semi-Angle of 100mrad 35

7.1.2 Previous Simulations Cropped to 50mrad 37

7.2 Training to Evaluate C1,2 in Ronchigrams Emulating the C3,0-Uncorrected STEM 38

7.2.1 Training with 85,000 Simulations 42

7.2.2 Training with 500,000 Ronchigrams 45

7.3 Training to Evaluate C1,2 in Ronchigrams Emulating the C3,0-Corrected STEM . . . 53
7.3.1 Training the Network Using Simulations . . . 53
7.3.2 Experimental Ronchigrams . . . 62

8 General Discussion and Conclusions **72**
8.1 Conclusion . . . 72
8.2 Future Work . . . 73

9 Ethical and Sustainability Considerations **75**

10 Project Management **78**
10.1 Initial Ultimate Objectives and Milestones . . . 78
10.2 How Early Results Matched Hypotheses . . . 79
10.3 Mitigating Setbacks in the Project . . . 79
10.4 Managing Resources . . . 80
10.5 Project Management Learning Experiences . . . 81

11 Risk Assessment **89**

12 Bibliography **90**

13 Appendix **98**

3 Glossary

ADF: Annular Dark-Field

ANN: Artificial Neural Network

BF: Bright-Field

β: convergence semi-angle

CNN: Convolutional Neural Network

CPU: Central Processing Unit

CTEM: Conventional Transmission Electron Microcope/Microscopy

CTF: Contrast Transfer Function

CX,Y: uppercase denotes the aberration of order X and rotational symmetry order Y in Krivanek notation; its magnitude is cX,Y and its angle is phiX,Y

LAADF: Low-Angle Annular Dark Field HAADF: High-Angle Annular Dark Field FC: Fully-Connected

Ground truth(s): label(s), aberration value(s) in the case of this project, associated with each CNN input image

Inference: aberration measurement calculation from a Ronchigram by a CNN

Iteration: simultaneous prediction, by a CNN, of labels of a single batch of training images

Learning rate: a value controlling the size of an incrementation step in gradient descent

AE: Absolute Error

MAE: Mean Absolute Error

MAPE: Median Absolute Percentage Error

MSD: Mean Squared Difference

NaN: *Not a Number*, an unrepresentable numeric data type

numpy: a Python package containing mathematical functions that may be included in code

Order: unless otherwise stated, this is the n-value of an aberration in Krivanek notation[1]

px: number of pixels

Input space: a term for how much simulations in appearances can differ based on the ranges that can be sampled from to form them

ImageJ: a Java-based image processing program

STEM: Scanning Transmission Electron Microscope/Microscopy

θ: detector semi-angle, equal to half the side length of any simulation or in this thesis

[a, b): range between values a and b, where the square bracket indicates a is included but the curved bracket indicates b is excluded

Aberration in Krivanek Notation	Aberration Name
C1,0	Defocus
C1,2	2-fold astigmatism
C2,1	2nd-order axial coma
C2,3	3-fold astigmatism
C3,0	3rd-order spherical aberration
C3,2	Star aberration
C3,4	4-fold astigmatism
C4,1	4th-order axial coma
C4,3	3 lobe aberration
C4,5	4th-order 5-fold astigmatism
C5,0	5th-order spherical aberration
C5,2	5th-order 2-fold astigmatism
C5,4	5th-order 4-fold astigmatism
C5,6	5th-order 6-fold astigmatism

Table 1: Table naming aberration constants in Krivanek notation[2].

4 Introduction

4.1 Transmission Electron Microscopy

Since the early 1940s, Conventional Transmission Electron Microscopy (CTEM) has been attempted to attain sub-nanometre resolution for studies like cell biology, where nano-scale hierarchies demand fine-scale analysis. Even the most high-resolution visible-light microscopes are only capable of resolutions around 20nm[3], too low to adequately observe DNA, whose double helices are only around 10nm wide[4]. Such deficiencies necessitate resolutions provided by short wavelengths of electrons accelerated in TEM. From top to bottom, the CTEM (an example of which is depicted by Figure 1) contains:

1. An electron source;
2. A magnification system in which the objective lens ultimately focuses the electron beam onto an electron-transparent sample immersed into this lens' magnetic field; and
3. A projection system, including detectors generating images from beam-specimen interaction signals, like barely-scattered transmitted electrons when in BF mode.

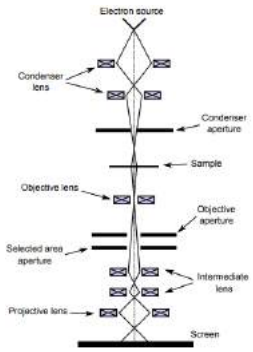


Figure 1: Schematic of a CTEM from[5], reuse permission granted by the author Naresh Marturi.

However, STEM differs from CTEM. While CTEM is advantageous when it is desired to record the contrast of all specimen pixels at once, STEM is advantageous in that it may collect individual signals of many different types simultaneously at each specimen pixel, these signals including BF, LAADF, HAADF and more. In the STEM, a field-emission electron gun and electron lenses (which ultimately converge the beam unlike in CTEM, where illumination is parallel) are arranged as in Figure 2. This allow the beam

to achieve resolutions even higher than 0.2nm[6]. In the STEM Figure 2 depicts, pre-specimen lenses demagnify the electron source to an atomic-scale probe on the specimen's surface, scanning coils translating the probe over the specimen in a raster. Signals one chooses to collect at each specimen pixel include aforementioned BF signals, and electrons diffracted by the specimen at high angles from the optic axis in ADF mode[7]. Objective lens aberrations dominate demagnification[7], given this lens's proximity to the specimen and probe-forming responsibility.

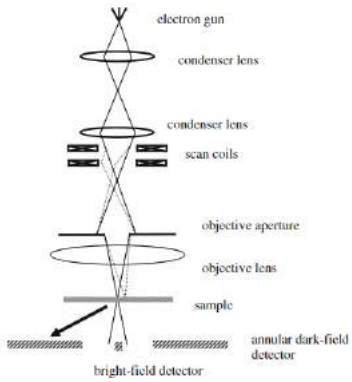


Figure 2: STEM schematic from[7].

Scherzer[8] showed aberrations are intrinsic to rotationally symmetric electromagnetic fields formed by lenses, the objective lens being most affected by aberrations due to dealing with the widest angular ranges of electron incidence. Thus, this lens is the most crucial in alignment for high resolution imaging in the CTEM[9], and also in STEM due to its equivalence to CTEM[7]. Among others, objective lens aberrations include:

1. Spherical aberrations from increasing electron focus by the field with distance of the electron path from the optic axis;

2. Comatic aberrations from lacking collinearity between optic axis and beam axis; and

3. Astigmatic aberrations from lens design errors, like those permitting magnetically non-uniform lens material[10], causing beams in perpendicular planes with dissimilar focal lengths.

Aberration evaluation permits efficient correction and resolution increase. In STEM, evaluation involves analysis of the electron Ronchigram.

4.2 The Ronchigram and its Aberrations

For objective lens aberration measurement in STEM, the electron beam is typically converged by the lens onto a thin, amorphous specimen whose near-random atomic potentials form of a diffraction pattern. Features of this are influenced by the aberrations to extents governed by their magnitudes and–for those like C1,2 which aren't axially symmetric–angles. The Ronchigram isn't typically used for alignment in the CTEM where beam convergence onto the specimen is lacking[11]. However, the Ronchigram is vital for this in STEM[1]. Approximating the aforementioned specimen type as a noise grating with atomic potentials described by $V_{noise}(x)$ permits the following quantities to detail the Ronchigram[1]:

- Intensity of electron probability density on the diffraction plane, describing the Ronchigram itself (Equation 1);
- Transmission wavefunction, describing the specimen-diffracted beam (Equation 2);
- Probe wavefunction, describing the probe before specimen interaction (Equation 3); and
- Lens aberration function, describing its deviation from ideality (Equation 4).

The aforementioned equations are as follows:

$$g(\alpha, \phi) = |IFT\{\psi_t(\mathbf{x})\}|^2 \tag{1}$$

Here, α and ϕ are the semi-angle (from the optic axis, which is normal to the objective lens plane) and azimuthal angle (about the optic axis) respectively of electrons incident at each Ronchigram point. ψ_t(x) is the transmission wavefunction in real space x:

$$\psi_t(x) = \psi_p(\alpha, \phi) \cdot \exp\{-i\frac{\pi}{4} \cdot \sigma V_{noise}(x)\} \tag{2}$$

Here, σ is the interaction parameter, equal to $2\pi m_e \lambda / h^2$ where m_e is relativistic electron mass, e is electron charge and λ is electron wavelength. The exponent shown represents random phases added to the beam by the Eikonal approximation[1], i's multiplier being the beam phase shift from the specimen[12]. $\psi_p(\alpha, \phi)$ is the probe wavefunction:

$$\psi_p(\alpha, \phi) = FT\{e^{-i\chi(\alpha,\phi)}\} \tag{3}$$

Here, $\chi(\alpha, \phi)$ is the lens aberration function:

$$\chi(\alpha, \phi) = \frac{2\pi}{\lambda} \cdot \Sigma_{n,m} \frac{c_{n,m}\alpha^{n+1}cos(m(\phi - \phi_{n,m}))}{n+1} \tag{4}$$

Here, α and ϕ mean the same as in Equation 1 except now being for the converged beam. $c_{n,m}$ is the magnitude of the geometric lens aberration of order n and rotational symmetry order m in Krivanek notation, and $\phi_{n,m}$ is the aberration's angle[13]; note that arbitrarily redundant angles are separated by 2π/m from any quoted $\phi_{n,m}$.

Figure 3 explains the Fourier transforms in Equations 1 and 3. These transforms can model lenses[14], so Equation 3 includes one for probe formation by the objective lens, and an inverse Fourier transform is applied to the transmission wavefunction due to the diffracted beam's divergent nature.

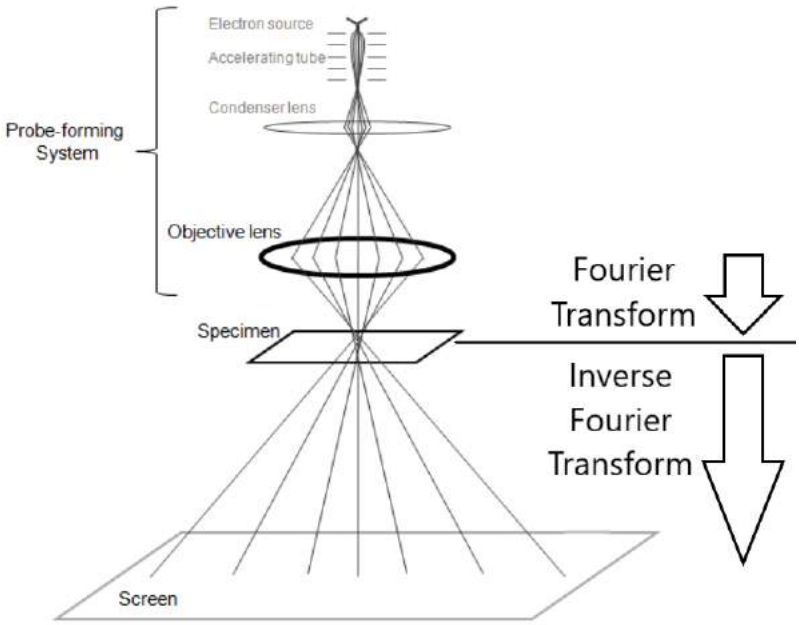


Figure 3: Schematic from JEOL of Ronchigram acquisition in STEM[15], edited with Fourier transforms.

Figure 4 is a simulated Ronchigram whose higher-order aberrations cause fringes surrounding its central region. For good alignment, lower-order aberrations are minimal and uniformity exists in the central region. This region is high in magnification, a factor varying throughout the Ronchigram and depending on the aberration phase's second derivative[16]; high magnification means a fine probe, and thus high resolution in STEM.

However, with lower-order aberrations, the central region becomes non-uniform.

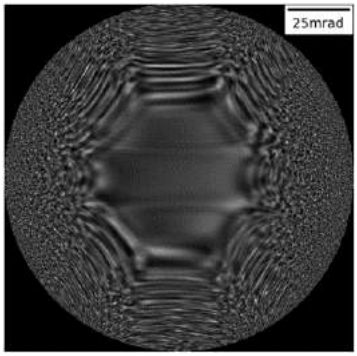


Figure 4: Simulation with only 3rd-5th order aberrations (with magnitudes and angles equal to midpoints of their ranges in Table 4).

For example, Figure 5 demonstrates the combined effect of C1,2 and C1,0, the former uni-directionally streaking the Ronchigram (see Figures 5 and 6) and the latter accentuating this pattern (compare Figures 5 and 7). Figures 8 and 9 demonstrate the effect of C2,1 on the Ronchigram shown in Figure 4, shifting the Ronchigram's centre. Finally, instead introducing C2,3 triangulates Figure 4 as in Figures 10 and 11.

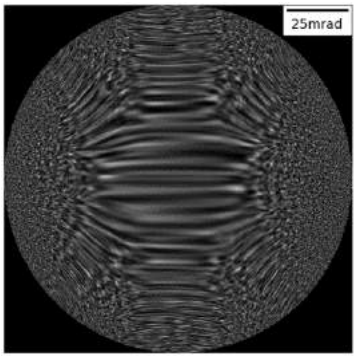


Figure 5: Figure 4 with c1,0 = 50nm, c1,2 = 50nm and ϕ1,2 = $\pi/2$ rad.

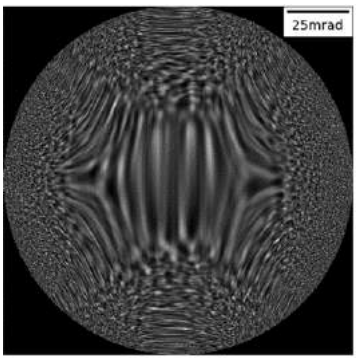


Figure 6: Figure 5 with ϕ1,2 = 0.

If uncontrolled, aberrations in the slightly dynamic objective lens drift, distorting the

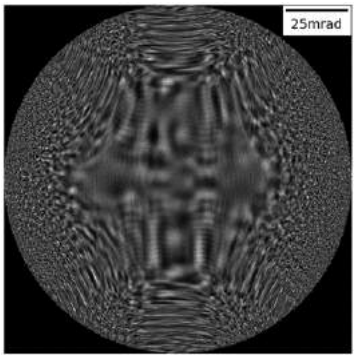


Figure 7: Figure 5 with c1,0 = 0.

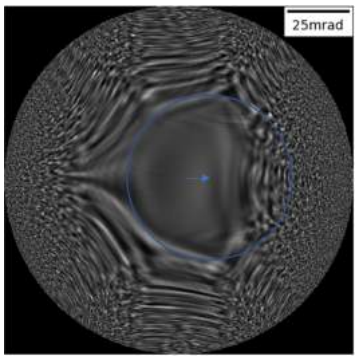


Figure 8: Figure 4 with c2,1 = 800nm and ϕ2,1 = π rad.

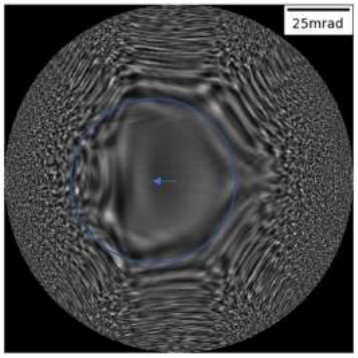


Figure 9: Figure 8 with ϕ2,1 = 0.

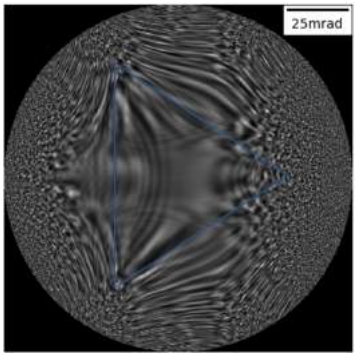


Figure 10: Figure 4 with c2,3 = 800nm and ϕ2,3 = $\pi/3$ rad.

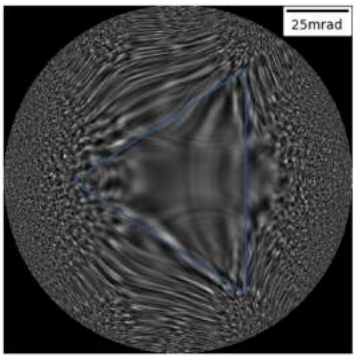


Figure 11: Figure 10 with $\phi 2{,}3 = 0$.

Ronchigram, so too long after aberration correction, resolutions achieved begin to decline. Lower-order aberrations drift quicker, as higher-order ones like C2,3 partially decompose into lower-order ones when they drift, further drifting lower-order aberrations. So, C1,0 and C1,2 tend to drift most severely given their low orders, the latter drifting in angle as well as magnitude while C1,0 is axially symmetric. For these reasons, priority in this project was given to enabling CNN performance on C1,2. Moreover, image shift (mentioned in Table 1) was omitted from simulations and evaluation because the Ronchigram should look near-identical whether generated using one point of the specimen or another a few nanometres away, making it negligible here.

4.3 Machine Learning for Aberration Measurement

As an aberration's characteristic pattern becomes more visible throughout a sequence of Ronchigrams where the corresponding magnitude increases, it may be learnt that magnitudes should be higher where this pattern is more visible. Similarly, as the pattern rotates throughout these Ronchigrams while the corresponding angle changes, pattern orientation may be learnt to determine angle value. Equations 1 to 4 mean each Ronchigram point is superimposed by patterns of all aberration present so, since such patterns can determine aberration values, a machine can learn values depicted by a Ronchigram if it *sees* enough pixels in enough Ronchigrams along with knowledge of values featured; a CNN is a machine with the potential to evaluate accurately even through the random noise present in the Ronchigram.

After training, a CNN (see Figure 12) can take arrays of image data (like Ronchigram pixels) and predict labels (like aberration values) associated with each. Within it are convolutional layers in which *filters*, each searching for a different feature, slide across the

image, measuring the presence of each feature, this feature becoming increasingly complex in deeper layers. Regarding the Ronchigram, while filters at earlier layers may learn to measure basic features unidirectional streaking from C1,2, at deeper layers, more complex features from the superposition of aberration patterns may be searched for.

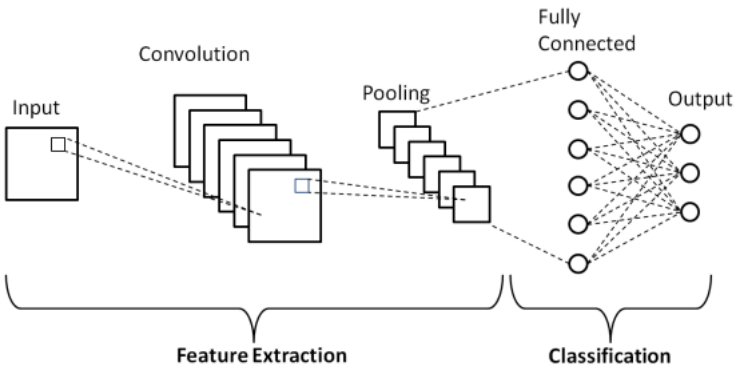


Figure 12: CNN schematic from[17] displaying input, feature extraction and classification. Reuse permission granted by Van Hiep Phung.

A filter is a matrix of *weights* convolved over image pixel groups, at each location producing a segment of an *activation map* via a dot product between the filter and the pixel values in the case of a *linear classifier*. After convolution, the activation map depicts quantities of the sought-after pattern, said maps inter-dispersed with:

1. *Pooling layers*, down-sampling them to compute further operations on summaries of the image, making the CNN more robust to varying locations of features in an image[18], essential for measuring an aberration given its effect on the entirety of a Ronchigram; and

2. *Activation functions*, which transform activation maps into metrics of pattern predominance in an image, which are transmitted through neurons in the CNN, leading to label prediction at the FC layer in the case of EfficientNet[17] (see Section 6.2).

In *training*, each image's predicted label is compared with the ground truth, this comparison evaluated by a *loss* metric, typically selected such that its value is greater for less close comparisons, like for MSE loss. Therefore, in *gradient descent*, loss *minima* are sought by, back-propagating losses calculated by the CNN towards the weights that generated the prediction, gradient descent updating the weights by an extent influenced by this loss's value, intending to travel downhill towards a *loss minimum*. Here, weights

predict values with low loss and high closeness to ground truth. Alongside training is *validation* where, after each epoch for example, the CNN predicts ground truths of *validation* images. These aren't shown to the CNN during gradient descent so don't affect learning– they instead measure CNN performance and ability to generalise learning to data outside of training. This can be quantified by metrics like MSE measuring CNN performance on the validation set.

By searching for loss minima in gradient descent while ensuring generalisation to Ronchigrams outside training simulations, a CNN may learn to evaluate aberrations in the objective lens using which these Ronchigrams are formed.

4.4 Engineering Context

Short wavelengths in TEM permit analysis of nanometre DNA structures (like double helices, see Figure 13) in cancer research. Specifically, the average human nucleotide is only 6 long[19], necessitating high-resolution forms of TEM like STEM, which can achieve resolutions of 0.2nm[20] and more and has been used to study EGFR receptors in COS7 cells, which are instrumental in various cancer[21].

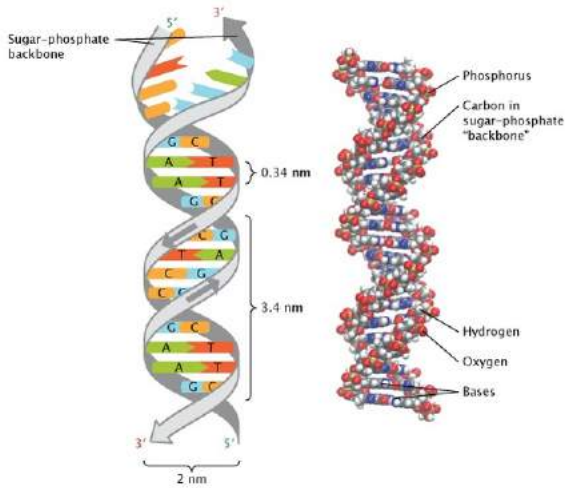


Figure 13: Double helix drawing from [22].

Correcting aberrations after measurement from the Ronchigram allows the STEM to be calibrated as near its maximum resolution as possible for reliable use in applications like aforementioned. However, maintaining this resolution throughout investigation necessitates real-time measurement, lest aberration drift cause beam misalignment and compromise resolution. Therefore, measurements must be fast. Machine learning on the

can indeed be sufficiently fast, with a 219ms inference time per Ronchigram by this project's CNN on a typical CPU, and more importantly has the potential to remove the need for time-consuming adjustments like raising electron illuminations; such an action also risks sample damage by the probe, given the high electron current it focusses on a small area of the specimen as in Figure 22. See Figure 14 for an example of such damage, damage being especially detrimental if measurements are to be made in-situ to reduce time consumption, and such damage limits discoverable information. Given the limitations of methods like auto-correlation in this regard, training CNNs for aberration measurement from Ronchigrams is desirable, especially given their successes in high-noise images (like low-illumination Ronchigrams are) already[23].

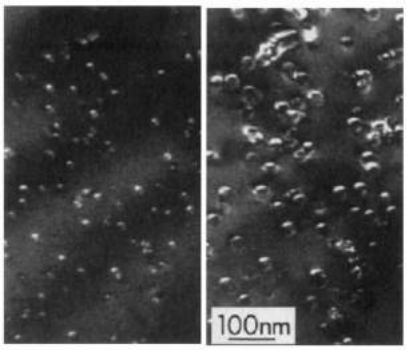


Figure 14: Damage (bright bubble-like regions) in quartz by 125 keV bombarding electrons, time increasing from left to right. Figure from [24].

The practical CNN-for-aberration-measurement workflow would simply be:

1. Ronchigram acquisition; and

2. Ronchigram image transmission to a CPU for inference by the CNN

The conformability of these steps would permit using other measurement methods at high illuminations if desired while switching to the CNN at low illuminations, and a new CPU would rarely be warranted, especially given that lower-memory networks can always be used if trained with help of this thesis' results. Finally, especially since a CNN has already been shown to measure C2,1 in Ronchigrams from varying specimen types[25], meaning no extra material must be bought for measurement, the negligible material and apparatus needed mean easy commercialisation of this method, especially that there are only several thousand electron microscopes existing and developed annually[26], many fewer of which are STEMs.

Despite results presented for the recognition of other aberrations, the findings of this report are ultimately limited to evaluation of C1,2 in the C3,0-corrected STEM at high illuminations for thin, amorphous specimens. However, due to drift and instability mentioned in Section 4.2 (hence the measurement of C1,2 in image correctors[27]), it was imperative to prioritise this aberrations for proof-of-concept, especially as it has the same rotational symmetry as many higher order ones. Practically, while C2,1 and C2,3 must be corrected only every 2/3 hours, C1,2 requires more frequent checking[28]. Typically, c1,2 should be below 1nm for optical resolution in STEM[12].

5 Literature Review

Common aberration measurement methods from Ronchigrams are quick and accurate at high illuminations, yet fail at low ones; so, after ensuring CNN performance at high illuminations, the aims were to continue success in Ronchigrams featuring decreased electron currents and acquisition times. A lesser aim was to extend success to Ronchigrams from specimens other than thin, amorphous ones, to ensure ubiquity of real-time use.

One commonplace method of aberration measurement from Ronchigrams is using Zemlin tableaus, whose slow implementation (requiring acquisition of 10 sets of STEM images at multiple probe tilts, taking 30-60s[12]) makes it inadequate for real-time use–for reference, see Figure 15 for examples of the results of such tilts, although for 16 tilts in the TEM rather than 10. According to Hawkes[29], Zemlin tableaus[30] had been the most used tool for aberration measurement from Ronchigrams.

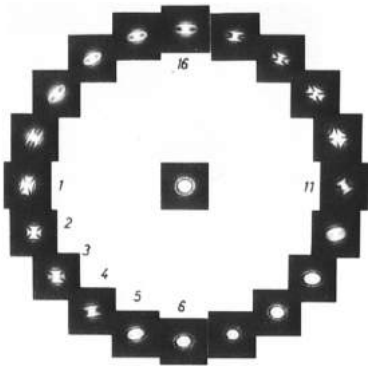


Figure 15: Zemlin tableau containing 16 diffractograms from a TEM. Figure from [30].

However, also according to Hawkes[29], phase correlation[31][32] (which is drawn from in JEOL's auto-correlation function[2]) was becoming advantageous regarding Ronchigrams from crystalline as well as amorphous specimens. For reference, the phase correlation function is:

$$PCF(x) = FT^{-1}[F(k)\frac{c_1^* c_2}{|c_1^* c_2|}] \tag{5}$$

Here, F(k) suppresses noise[29]; decreasing Ronchigram acquisition illuminations means lowering image signal-to-noise ratios, increasing detector noise, so suppression is suboptimal when preventing sample damage by the probe and assisting real-time aberration measurement. Comparatively, although proof is lacking, CNNs could overtake phase correlation in measurement at high detector noise levels after training with high-noise

simulations. Furthermore, in limited work done so far, CNNs have retained phase correlation's ability to evaluate aberrations in Ronchigrams acquired using materials beyond thin, amorphous samples, at least for C2,1[25].

Aberration measurement and correction methodologies using machine learning without Ronchigrams include reinforcement learning, like in spherical aberration correction in the STEM[33]. However, quotes no times taken, meaning feasibility in real-time aberration measurement is unproven. Besides methods without Ronchigrams, successful applications utilising them are also scarce. However, machine learning has ascertained quantities other than aberration magnitudes and angles from the Ronchigram for resolution maximisation, like optimal convergence angles in the STEM as done by Schnitzer et al.[13]. Here, for optimal convergence angle selection, a CNN is trained to learn recognise aberration functions (see Equation 4) from the Ronchigram–see Figure 16 for this process. However, learning the aberration function does not equate to evaluation specific angles and magnitudes needed to correct aberrations. Moreover, this paper does not present success for low illuminations. In contrast, one of the closest attempts to the current project is a CNN by JEOL to indeed recognise aberration magnitudes and angles, specifically those of C2,1[25]. Although claims of success in multiple specimen types, results are presented only for C2,1 while the current project initially aimed at several low-order aberrations; besides, in this project, C1,2 was prioritised, generally being more unstable than C1,2. Furthermore, the precision quoted in the aforementioned paper for c2,1 and ϕ2,1 is 20%. This is acceptable for quick, repetitive correction, where correction is faster than C2,1 drift. However, 20% could still be beaten, and [25] again avoids mentioning success at low illuminations.

The work nearest this project was released and discovered only late into this project; Bertoni et al.[12] used an ANN rather than a CNN (although it is based on the VGG-16 type CNN[34]) to evaluate all of C1,0, C1,2, C2,1, C2,3 and C3,0 from Ronchigrams. They use a more extensive method to process Ronchigrams before network input, by segmenting the Ronchigrams as inspired by used of the CTF[35]–see Figure **??** for this process.

Promising results are quoted, including MAEs of, for each aberration value, only around 5% of the maximum value in training simulations, corresponding to resolutions beating 100pm. In the aforementioned paper, similar ranges are used for C1,2 as in Table 4, although without negative values for the latter. Bertoni et al.'s method is also fast,

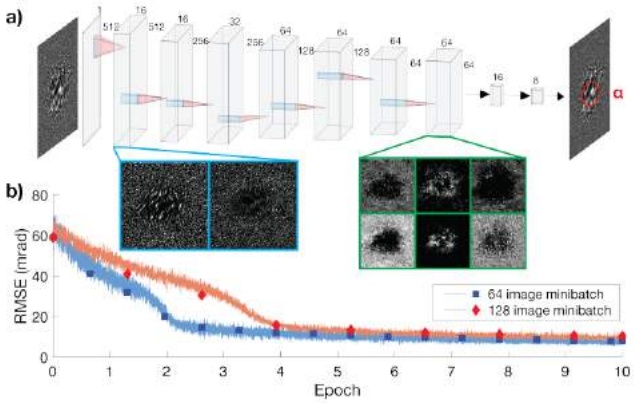


Figure 16: (a) is the CNN schematic used and (b) its success in evaluating optimal convergence angles for two different batch sizes (solid lines). Markers are validation errors. Figure from [13].

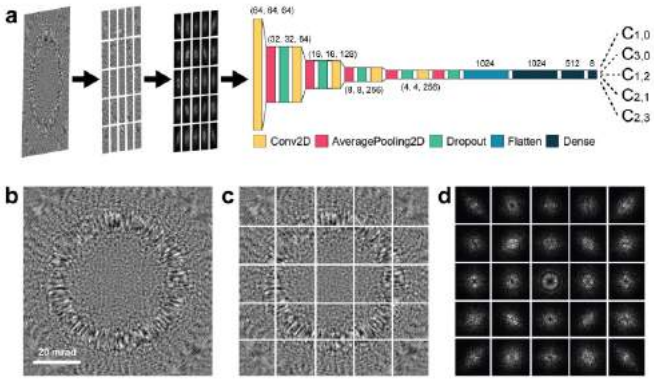


Figure 17: (a) is the ANN, (d) the input to the network generated from the Ronchigram in (b). Figure from [12].

with 150ms necessary for aberration diagnosis from one Ronchigram on a standard PC (alongside a 1-2s Ronchigram acquisition time in practice); this Part II project aimed to acquire one Ronchigram per measurement, so similarly low measurement times were expected. However, Bertoni et al. admitted possible limitations of success being shot noise (Poisson noise) and camera noise, simulation of which no mention is made of. As well as initially aiming to ensure success at low illuminations, this Part II thesis at least discusses simulating Poisson noise in Ronchigrams (see Section 6.1.2), where electron currents and Ronchigram acquisition times are randomly sampled (using a uniform probability distribution, like all random sampling in this thesis unless otherwise stated) from ranges to do this. Finally, their paper only present results for thin, amorphous carbon, which this Part II project eventually aimed to go beyond. Therefore, regarding noise and specimen types, the initial aims (and some details) of the current project are distinguished from the work in [12].

To build upon the literature, this Part II project aimed to, ultimately, train a CNN to evaluate aberration magnitudes and angles from Ronchigrams under decreasing electron illuminations and varying specimen types to achieve real-time aberration measurement. This was to reduce specimen damage by the beam by removing needs to tune electron dose when measuring in-situ. These freedoms were to help to close the gap between microscopists and real-time resolution maximisation in STEM, enhancing the results and rates of scientific investigation. This project prioritised performance on the unstable, axially symmetric C1,2 in the C3,0-corrected STEM and aimed to inform further training. Finally, this aberration measurement method was to necessitate little extraneous experimental apparatus, meaning ubiquity in its potential for use by microscopists with varying resources available.

6 Methodology

6.1 Ronchigram Simulations

6.1.1 Simulations Before Detector Noise Application

Ronchigrams were simulated for training and validation in Python as 1024x1024 numpy arrays, emulating a thin, amorphous specimen in the STEM at 300kV electron accelerations. As well as using Equations 1 to 4 for this, projected potential $V_{noise}(x)$ was specially formulated. To simulate this, an over-sampled white noise grating was used, given past successes in roughly approximating the aforementioned specimen type[1][13]. Multi-slice simulation packages can do this accurately but add unwarranted computational overhead, especially through accounting for the effects of multiple scattering[13] which is negligible for this specimen type, meaning such packages weren't necessary at this stage of the project.

Regarding $V_{noise}(x)$, the multiplier of -i in Equation 2 had an expectation value of $\pi/8$ chosen as done by Schnitzer et al.[36], acceptable given its closeness to estimates from the weak phase object approximation for a σ of $0.0065V^{-1}nm^{-1}$ with a 300kV beam for a specimen thickness of around 10nm and a mean inner potential of 10.7V for amorphous carbon[37][38]. Experimental Ronchigram acquisition was done using a similar specimen (see Section 6.5).

Finally, a condenser aperture is applied to the simulation by simulating a mask setting values of all pixels of the simulation outside of the circle touching all borders of the simulation to zero. Thus, convergence semi-angle was simulated to have the same value as detector semi-angle.

6.1.2 Detector Noise Application to Simulations

The only experimental artefact initially emulated in the simulations was detector Poisson noise, as the primary aim was to ensure CNN performance over varying illuminations, for which Poisson noise was considered most relevant since the detection of individual electrons may be considered independent events following a random temporal distribution[39]. Therefore, chromatic aberration, impulse noise from dead pixels in the detector[40] and Gaussian circuit noise were excluded from simulations, given their negligibility relative to high Poisson noises eventually to be simulated, and Poisson noise predominates image

noise at high illuminations anyway[39]. However, it was planned to include these artefacts later based on CNN performance on experimental Ronchigrams, in the fashion of the incremental development approach detailed in Section 6.4.

Ronchigram acquisition electron currents (I) and times (t) were randomly chosen from 100pA-1nA and 0.1s-1s (oft used in experimentation, according to Dr. Huang, making results in Section 7 comparable to auto-correlation, which measures aberration values at high illuminations), meaning low but noticeably varying Poisson noise (see Figures 18 and 19). Simulations without detector noise then had it applied using the Poisson distribution displayed in Equation 6.

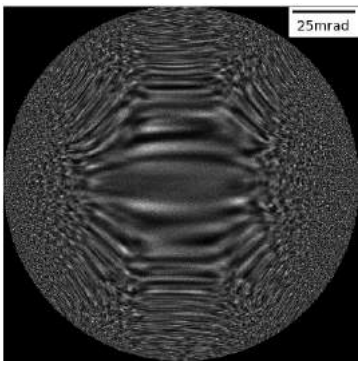


Figure 18: Low-illumination simulation. I = 100pA and t = 0.1s. Aberration values are midpoints in Table 4.

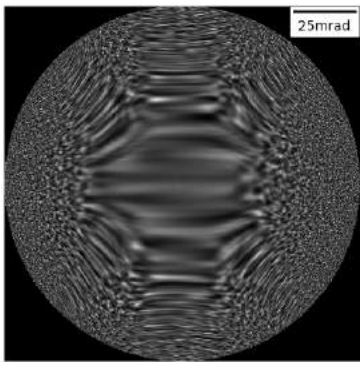


Figure 19: High-illumination simulation. Figure 18 with I = 1nA and t = 1s.

$$f(k;\lambda) = \frac{\lambda^k e^{-\lambda}}{k!} \tag{6}$$

Here, f(k) is effectively the probability of k events occurring per unit time, λ the mean number of events per unit time. To ensure events simulated were electron counts at a given pixel on the detector with alongside acquisition current I and time t, Equation 7

was derived for the λ array:

$$\lambda(\alpha, \phi) = \frac{I \cdot t \cdot b \cdot g(\alpha, \phi) \cdot \Sigma g(\alpha, \phi)}{e} \tag{7}$$

Here, e is electron charge, $g(\alpha, \phi)$ the array of electron probability densities at each pixel as in Equation 1, and the sum is of this quantity across the detector. b is the proportion of *quoted experimental current* reaching the detector, approximated as 1.

6.1.3 The Probe

In this project's simulation code (see Section 13), the real-space probe in the specimen plane (see Figure 3) before interaction with the specimen may be depicted using the probe wavefunction. Figure 20 depicts an unrealistically fine probe, a low c1,0 present alone. Here, the probe is well-encapsulated by the borders of this scanning step, simulating good sampling, where the probe doesn't cause blurring in STEM images due to a lack of overlap with adjacent specimen pixels (the probe is translated over the STEM in a raster, as in Section 4.1)–for realistic simulations, it's important they depict good sampling. Equation 8 is for the length of each specimen pixel imaged in Figure 20. In this equation, λ is wavelength and β convergence semi-angle in radians, equal to detector semi-angle as explained in Section 6.1.1.

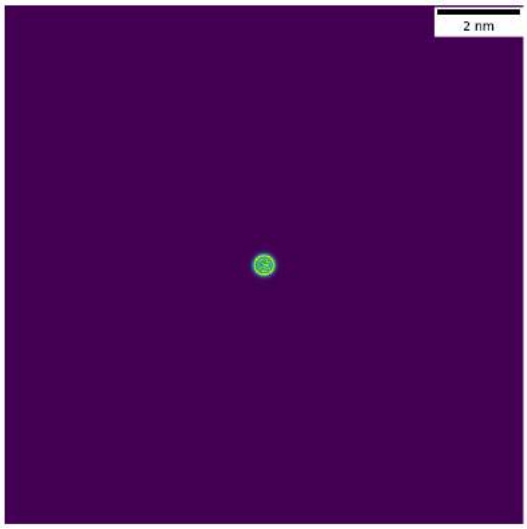


Figure 20: Probe simulation with c1,0 = 3.81nm, 1024x1024 px, β = 80mrad and dx = 0.123 Å.

$$dx = \frac{\lambda}{2\beta} \tag{8}$$

Other 1st-5th order aberrations magnitudes distort the probe into Figure 21, although it remains within image borders, meaning no oversampling. Introducing corresponding angles distorts Figure 21 into Figure 22, where sampling is also good. The latter features average aberration values in C3,0-corrected STEM simulations, meaning training in Section 7.3.1 would have approximated realistic sampling had an 80mrad convergence semi-angle been used. However, for reasons in Section 7.1.2, Ronchigrams were simulated there to 180mrad before cropping to 80mrad, so the average simulation in Section 7.3 may be a little oversampled (see Figure 23).

Because of a lack of extensive knowledge of the probe and experimental experience, the potential of this issue was not realised until late into the project. However, Figures 58 and 59 later demonstrate CNN recognition of trends in simulations simply made to 80mrad, although further training is warranted.

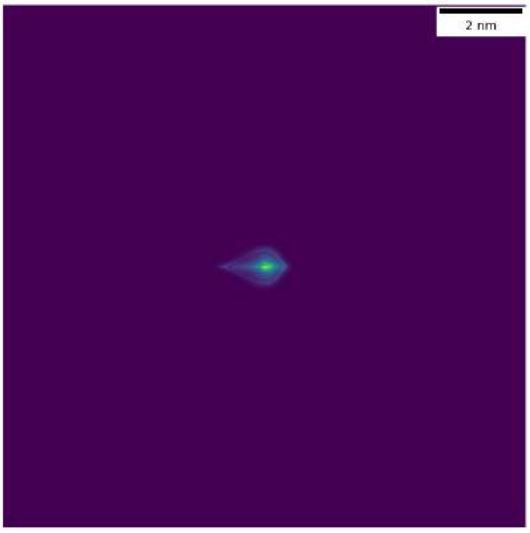


Figure 21: Figure 20 with aberration magnitudes equal to Table 6's midpoints but angles zero.

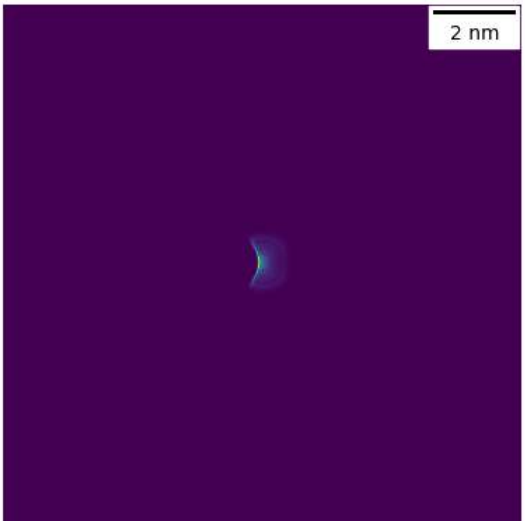


Figure 22: Figure 21 with aberration angles equal to Table 6's midpoints and a β of 80mrad.

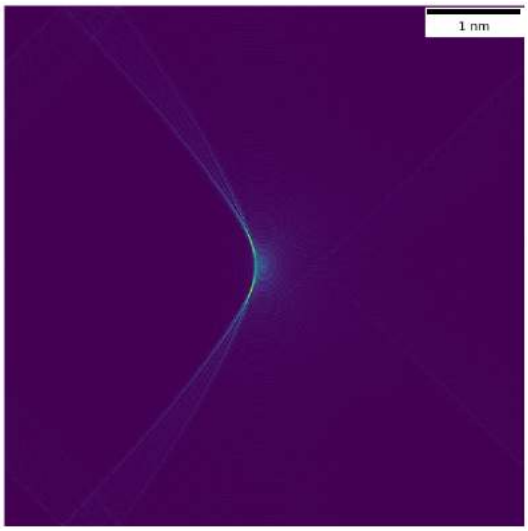


Figure 23: Figure 22 with a β of 180mrad.

6.2 Machine Learning Tools

To facilitate the machine learning, functions from popular open-source machine learning framework *PyTorch* were used, including:

1. A stochastic gradient decent optimiser[41], which increments the model weights in search of loss minima mentioned in section 4.3, Nesterov momentum and weight decay, of 0.9 and 10^{-3}, included to aid this given successes in EfficientNet implementations in [42];

2. A data loader[43] for loading batched Ronchigrams for training and testing the CNN; and

3. TensorBoard[44], to plot training and validation loss curves for training progress visualisation

The EfficientNet series[45] of CNN was chosen for this project, given its thorough documentation, competitive performance on ImageNet[46] (on which high accuracies were achieved in multi-class classification) and its past use in electron microscopy, in biological imaging[47]. In designing any EfficientNet model, a baseline CNN (EfficientNet-B0, see Table 1 of [45]) is upscaled by increasing network width (number of nodes in a convolutional layer), depth (number of layers) and resolution (number of pixels in an input image) using a fixed set of scaling factors to uniformly scale these properties–see Figure 24. Besides aforementioned EfficientNet success validation, this process should maximise accuracy because, for higher-resolution inputs, more layers are needed to increase the receptive field and more channels are needed to capture finer features[45].

The version used from this series wasn't expected to be an aberration evaluation accuracy bottleneck (data pre-processing was), so the selection criteria was ensuring the CNN could fit on the GPU (used for its ability to undertake parallel computing to support parallel CNN operations like matrix multiplication), a GeForce RTX 2080 TI, while permitting large enough batches (that are loaded one by one) for optimal training speeds. EfficientNet-B2 was the most developed version fitting this criteria.

Data augmentation using transforms is oft done to proliferate data and show a CNN variations of it in training for the same ground truth, to develop robustness. Such transforms may include *random affine*[48], which distorts images while keeping centres in-

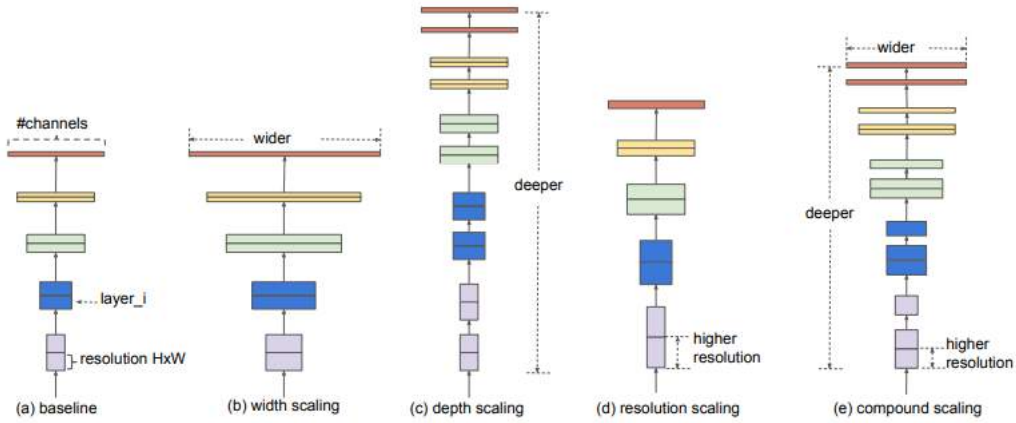


Figure 24: Scaling from baseline (a), with width, depth and resolution scaled individually in (b), (c) and (d) respectively, (e) depicting compound scaling. Figure from [45], reuse permission granted by Tan Mingxang.

variant. However, a given set of aberration constants has a unique contribution to the Ronchigram via Equation 4, so distortions may lead to an image the ground truth doesn't represent. Therefore, distorting transforms were avoided. However, acceptable transforms used in order included:

1. *ToTensor*[49], to convert numpy simulations to PyTorch Tensors for use in PyTorch functions;

2. *CenterCrop*[50], to select Ronchigram areas to train and test the CNN with;

3. *Resize* with *bicubic interpolation*[51], such that resolution was optimal for EfficientNet-B2 (260x260 px); and

4. *Normalize*[52] using the mean and standard deviation of the data, for pixel values optimal for CNN processing.

Bicubic interpolation and image normalisation had already achieved success in EfficientNet implementations[42], hence their use. Figures 25 and 26 compare a simulation before and after pre-processing, the latter depicting the version shown to the network.

Finally, ground truths were also normalised: for an aberration constant sampled from range [0, b) to generate a set of training or testing Ronchigrams, division of each value of said constant in the set by b was done such that ground truths were in the range [0, 1), a range of values unlikely to cause gradients in gradient descent that are too high for handling by PyTorch functions and thus lead to infinite values and NaNs that halt

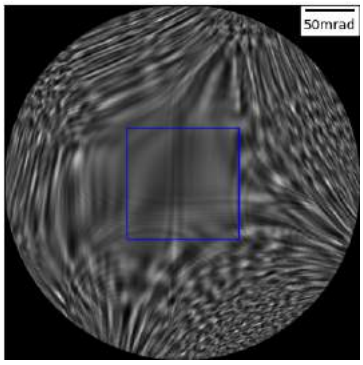


Figure 25: Simulation using parameters returned for Experimental Ronchigram 3 mentioned in Table 7. The box is the centre-crop.

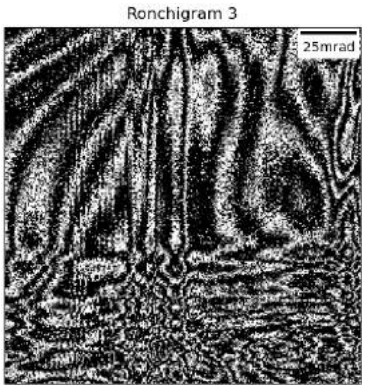


Figure 26: Figure 25 pre-processed.

training. However, the consistency of normalisation in both training and testing meant that CNNs trained on a range of aberration values [0, b), and subsequently learning to predict between 0 and 1, could still return constants when tested on simulations featuring that same range if predictions were subsequently multiplied by b.

6.3 Network Testing Methods

Two novel pipelines were generated to evaluate CNN performance on Ronchigrams: the *trend checker* and the *comparison checker*.

The first method takes a CNN and checks if it can predict a linear increase in an aberration value between simulations. Here, 1,000 simulations featuring this trend are generated, with *other* values random between selected ranges (typically those used for training), inference is done by the CNN, and a *trend graph* of predictions vs ground truth is plotted (see Figure 43 for example). This is a good test because, even without high evaluation accuracy, it is practically promising if a CNN can recognise where a value is bigger or smaller, such that this value can be shifted in the correct direction in correction,

the randomness of *other* values practicability in diverse experimental cases.

The second method shows a Ronchigram to the CNN, aberration value(s) are inferred and, alongside all non-inferred values and parameters like current and acquisition time, uses these predictions to plot a new simulation, ideally identical to the test Ronchigram. This gives a visual representation of CNN ability and also depicts multiple aberration values are simultaneously predicted successfully. See Figure 45 for an example. However, it must be stated that the random effects of projected potential were not able to be recreated.

6.4 Incremental Development

Training was incrementally developed, beginning with less challenging recognition tasks and evolving towards more realistic Ronchigrams. This was because it would be easier to identify causes of struggles for the CNN as they were being introduced (struggles presented throughout Section 7). In contrast, starting with the most realistic of Ronchigrams would have involved many factors that could have been hindering CNN performance. Moreover, comparing CNN performance on more elementary Ronchigrams with that on more complex one was to give a holistic sense of the CNN's capabilities.

During incremental development, it was aimed to ensure the CNN could at least recognise trends in aberration constants in inference using the trend checker mentioned in Section 6.3. However, attempts to fine-tune CNN performance to primitive simulations were avoided due to a lack of time and necessity, given it was the later-introduced, more complex experimental Ronchigrams the CNN was to perform optimally with. Not all results of incremental development are introduced in Section 7, but the most crucial results that informed the evolving training methodology are.

6.5 Experimental Ronchigram Acquisition

To evaluate the CNN's practical performance, Dr. Huang acquired Ronchigrams using a thin (approximately 20nm), amorphous carbon film grid in JEOL's JEM-ARM300F2 GRAND ARM™2 Atomic Resolution Analytical Microscope, in STEM mode. A 300kV accelerating voltage was configured (as emulated in Section 6.1.1), two aberration correctors used to tune aberration values for the experimental Ronchigrams, and an Orius camera with a detector pixel size of 9um was used at a camera length of 8cm to collect

Ronchigrams containing 2672x2672 px. Although aberration values were tuned manually, the JEOL COSMO™ aberration correction software estimated aberration magnitudes and angles of 1st-5th order (those of even higher orders negligible) via auto-correlation, returning them alongside Ronchigrams with errors estimated for each value (see Figures 75 and 76 for resultant error bars). Ronchigram acquisition currents and times were varied throughout the experimental Ronchigram set, for evaluation of the CNN at various illuminations.

Unfortunately, there are limitations in the experimental Ronchigrams. First, when adjusting illuminations, there were difficulties maintaining the probe, resulting in subsets of Ronchigram where illuminations were changed but aberration values were hard to maintain, making effective controlled comparison of the CNN at varying illuminations difficult. Additionally, although of a similar scale, the specimen used was about twice the thickness of that for which projected potential was simulated (see Section 6.1.1). Moreover, sample damage and a contamination issue resulting in carbon deposition at the probe-illuminated areas increased specimen thickness, thereby further diverging specimen thickness from simulations–see Figure 27 for a sense of what this might look like. Finally, issues with calibration meant that Ronchigram dimensions were not returned, and that inaccurate electron currents were returned as percentages of their maxima during acquisition. Although acquisition was done months in advance of this thesis deadline and pipelines had already been set up to quickly process experimental Ronchigrams, it was not able to be redone to avoid these issues due to backlogging of the availability of the same STEM.

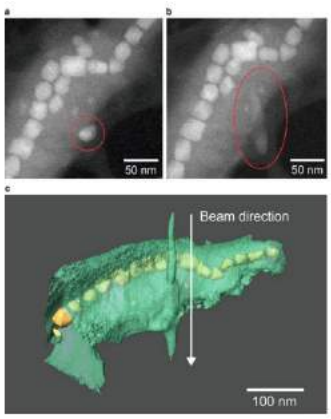


Figure 27: Example of contamination in STEM, the specimen in (a) left while the bright spot (presumably hydrocarbon) had time to elongate to its form in (b); reconstruction in (c) depicts protrusions from contamination parallel to beam direction. Figure from [53].

Because of these issues, evaluation using experimental Ronchigrams were an ambitious test of CNN robustness. Although still a worthy test, it must be remembered that poor results in Section 7.3.2 may under-represent the CNN's experimental capabilities.

7 Discussion of Results

7.1 Training the Network to Evaluate Aberration Constants in Simulations Featuring Only 1st and 2nd Order Aberrations

An early training run to determine if the CNN could ascertain aberration values at all involved simulating 85,000 Ronchigrams featuring only 1st and 2nd-order aberrations, randomly sampling values from ranges in Table 2. All aberration values were randomly sampled for training simulations rather than picking them from their ranges at fixed intervals, lest the CNN be biased to predicting values at these intervals alone in generalisation. Aberration magnitude maxima were quite arbitrarily chosen as multiples of 10nm, however, it was at least decided these would be smaller for c1,0 and c1,2 than c2,1 and c2,3. This was because, for equal magnitudes, higher order aberrations have smaller contributes to the lens aberration function and thus the Ronchigram (see Equation 4) than lower ones, given the increased exponent of semi-angles α (which are on the mrad scale in simulations, so greater exponentiation leads to a smaller value) and the increased denominator $n + 1$. Therefore, c2,1 and c2,3 were given higher maxima to offset this and attempt to form simulations with more balanced aberration patterns. Alongside these maxima, and in all simulations used for training, aberration angles were selected such that all possible contributions of an aberration to Equation 4 could be covered for a given magnitude. Finally, simulations were done to a detector semi-angle of 100mrad (with radial angles [-100nm, 100nm]). Figure 28 depicts the *average simulation* here, formed by midpoints in Table 2. In hindsight, C2,3 predominates here, appearing as hexagonal in the company of comparatively small 1st and 2nd-order aberrations and no higher-order aberrations (unlike in Figure 10, where it causes triangulation).

In each training run described in in the current Section 7.1, only 10 epochs of training were done, as detailed in Table 3. A maximum initial learning rate (referred to as maximal because training was split between three different groups of CNN parameters as in [42], in which learning rates were different percentages of this maximum) of 0.01 was selected. This was found to be sufficient in permitting training loss to decrease, as proven later by Figures 29 and 30. Additionally, the simulations were randomly split into two sets of a 14:3 ratio before training (as it was for all training done), the former for gradient descent and the latter for validation done after each epoch in this project to ascertain CNN ability

Parameter	Value or range
Total number of simulations associated with each training run	85,000
I range	[100pA, 1nA)
t range	[0.1s, 1s)
c1,0 range	[0, 10nm)
c1,2 range	[0, 10nm)
c2,1 range	[0, 1000nm)
c2,3 range	[0, 1000nm)
$\phi_{n,m}$ range	[0, $\frac{2\pi}{m}$rad)
θ	100mrad

Table 2: Parameters to simulate Ronchigrams used for training in Section 7.1. In the $\phi_{n,m}$ range, m is rotational symmetry order per aberration, drawn from a range including 0 but not $\frac{2\pi}{m}$rad because of the equivalence of the aforementioned maximum and minimum (see Equation 4).

to generalise learning from training. Finally, an MSE metric was used to measure both training and validation losses.

Parameter	Value
Maximum initial learning rate	0.01
Gamma by which learning rates were multiplied after each epoch	0.975
Loss criterion	MSE
Gradient descent : validation split of training simulations	14:3

Table 3: Parameters used to train the network at this training stage with simulations featuring only 1st and 2nd order aberrations.

In the next two sections, results are presented for the findings of the above process, in each section a different simulation extent shown to the network for training and testing. For this stage of the project, training was done using simulations from aberration angle ranges [0, π/m rad) rather than [0, 2π/m rad), so these results come from retraining done to make them comparable with results presented later in this section.

7.1.1 Simulations with a Detector Semi-Angle of 100mrad

The first of two examples of training runs using the aforementioned simulations was done after centre-cropping each simulation to the inner square of the condenser aperture, depicted by the blue box in Figure 28, since aberrations don't affect the mask in this simulation and thus the mask could not help train the CNN to evaluate them.

Figure 29 displays loss curves of the training run using these simulations, and overfitting is noticeable. Overfitting may occur when a network continues to learn to identify ground truths of the training data while showing comparatively inadequate success on

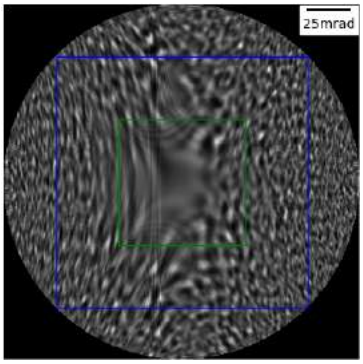


Figure 28: Average of simulations detailed by Table 2 with a 100mrad detector semi-angle. The blue box is a centre-crop for training in Section 7.1.1, the green box used in Section 7.1.2 to emulate aperture-cropped 50mrad Ronchigrams. Note: cropping refers to the simulations for which this one is an average, not this average simulation specifically.

validation data and thus in generalisation outside of training data–this is unacceptable since experimental aberrations can have infinitely more combinations than the training data (since constants can take continuous values), meaning learning the training data is insufficient. Here, overfitting is shown for both c2,3 and $\phi_{2,3}$ after separation of validation losses curves from the training loss curve after only around 5 epochs (11,000 iterations).

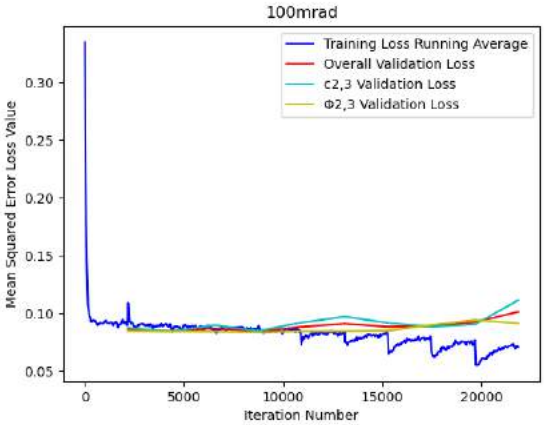


Figure 29: Loss curves for training the network to recognise C2,3 in simulations made using parameters in 2, cropped to the blue box in 28. Overall validation loss (the mean of that of c1,2 and ϕ1,2) is included because overall validation loss was used to automatically determine when early stopping was done. Training loss is meaned between those for c2,3 and ϕ2,3, and the running average of this mean is computed at each successive iteration. 2,190 iterations occurred per epoch.

Since overfitting occurred before validation losses were low enough for this training run to be progressed from, changes were to be made. A hypothesis for the cause of early overfitting was the inclusion of Ronchigram areas from highly-aberrated portions of the beam far from the optic axis (see noisy-looking off-centre areas in Figure 28). Although

aberration constants in some regard affect where these areas are, it is a complex task for a CNN to learn a rule for this to extend outside of training data, in comparison to simply learning C2,3 patterns, learning of which is currently obscured by attempts of the CNN to learn from pixels outside these patterns. In order to see if omitting highly-aberrated regions from training could improve generalisation, the amount of Ronchigram shown to the CNN was limited as in the following Section 7.1.2.

7.1.2 Previous Simulations Cropped to 50mrad

It was expected that cropping to the green box in Figure 28 would sufficiently exclude highly-aberrated portions of the Ronchigram from training and thus reduce overfitting. However, instead of simulating Ronchigrams detailed by Table 2 and then cropping to the inner square of the condenser aperture, the green box in Figure 28 was to approximate the effects of this due to advice that detector semi-angle was not typically varied quickly in experimentation and instead held to 70mrad-100mrad. Although lessons learnt in Section 6.1.3 mean the aforementioned cropping should not have been done, test Ronchigrams (for example in validation) were cropped in the same manner for consistency and thus inference success of the trained CNN.

Using crops done to the green box in Figure 28, the CNN was trained again with parameters in Table 3. Although Figure 30 demonstrates little change in learning ϕ2,3 compared to Figure 29, c2,3 no longer seems to overfit and instead continues to decline in validation loss–this implies such limitations of training to within the borders of the central region of the Ronchigram at least improved generalisable learning of aberration magnitude here. Furthermore, such improvements while extensive cropping permitted the use of such cropping to be continued.

In light of the above results, it was decided that further training runs would maintain extraneous cropping of simulations of presumed realistic simulation dimensions in order to emulate Ronchigrams of smaller detector semi-angles. More importantly, it was decided that crops would omit highly-aberrated Ronchigram areas from training and better expose characteristic aberration patterns in the central region instead.

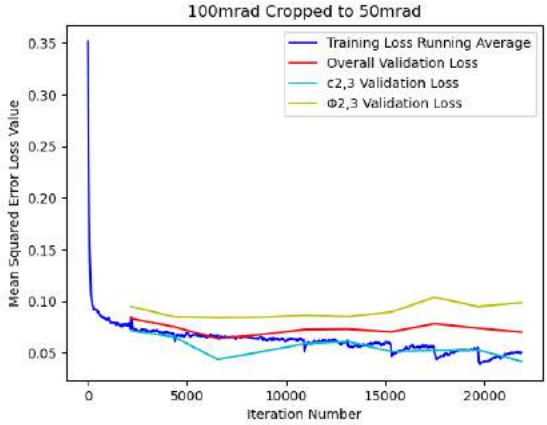


Figure 30: Loss curves for training done using crop as in the green box in Figure 28. 2,190 iterations occurred per epoch.

7.2 Training to Evaluate C1,2 in Ronchigrams Emulating the C3,0-Uncorrected STEM

An early attempt to train the CNN on experimentally realistic simulations was done using simulations based on aberration tolerances in the NION 100kV UltraSTEM, their magnitudes in Table 2 of [54]–here, c3,0 tolerance is 10.4um, a value typical for the c3,0-uncorrected STEM. These tolerances are reiterated as maxima of ranges in Table 4, except with 1st and 2nd order aberration maxima increased to higher multiples of 100nm in attempts to make them more visible, this being done as a first pass to test the CNN's capabilities in complex Ronchigrams. However, despite the aforementioned attempts, it was difficult to balance these aberrations, even using a similar scaling method to that in Section 7.1; C1,2 ended up predominating, as the uni-directional streaking in Figure 31 confirms. Therefore, it was at this stage that evaluation of C1,2 (prioritised for reasons mentioned in Sections 4.2 and 4.4) was begun. Alongside the aforementioned magnitudes, aberration angles were sampled from the range [0, $\frac{2\pi}{m}$rad) once again, as for previous simulations.

Ronchigrams in this section were simulated with a detector semi-angle of 70mrad and centre-cropped (to the green box in Figure 31, emulating an aperture-cropped 30mrad Ronchigram) once again, such that the only region shown to the network was that heavily influenced by c1,2 and ϕ1,2 in the average simulation–compare Figures 32 and 33, where c1,2 and ϕ1,2 are perturbed from the average simulation by 50%, to the average simulation in Figure 31.

Parameter	Value or range
I range	[100pA, 1nA)
t range	[0.1s, 1s)
c1,0 range	[0, 100nm)
c1,2 range	[0, 100nm)
c2,1 range	[0, 300nm)
c2,3 range	[0, 100nm)
c3,0 range	[0, 10.4um)
c3,2 range	[0, 10.4um)
c3,4 range	[0, 5.22um)
c4,1 range	[0, 0.1mm)
c4,3 range	[0, 0.1mm)
c4,5 range	[0, 0.1mm)
c5,0 range	[0, 10mm)
c5,2 range	[0, 10mm)
c5,4 range	[0, 10mm)
c5,6 range	[0, 10mm)
$\phi_{n,m}$ range	[0, $\frac{2\pi}{m}$rad)
Detector semi-angle	70mrad

Table 4: Parameters used to simulate C3,0-uncorrected-STEM Ronchigrams in Section 7.2. In the $\phi_{n,m}$ range, m is rotational symmetry order per aberration.

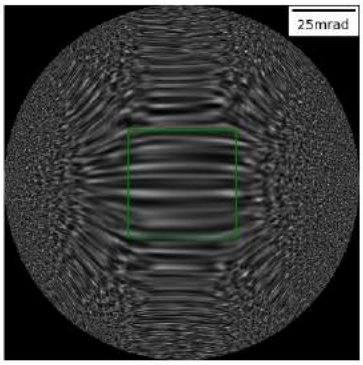


Figure 31: The average C3,0-uncorrected STEM simulation with the crop used for training indicated.

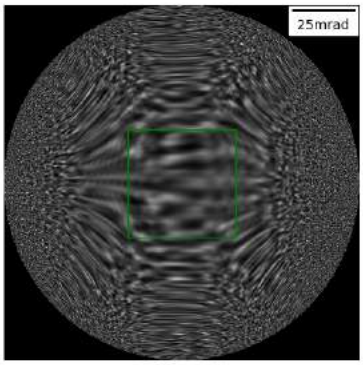


Figure 32: Figure 31 with c1,2 decreased by 50%.

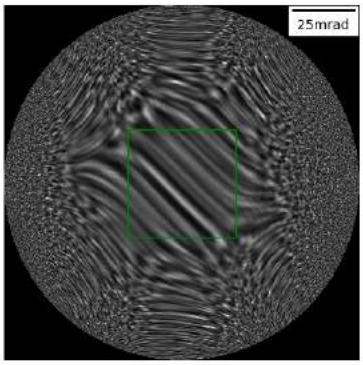


Figure 33: Figure 31 with ϕ1,2 increased by 50%.

Training using the aforementioned simulations was done with parameters in Table 5, similar to Table 3. However, the previous maximum initial learning rate of 0.01 was too high to decrease training losses here, potentially due to increased sensitivity of loss to changes in model weights in the comparatively complex simulations here. Such sensitivity necessitates smaller weight incrementation for downhill gradient descent, lest changes cause overshoot to higher points on opposing loss hills, given their increased proximity among the more complex loss peaks–see Figure 34. The decreased maximum learning rate to 0.001 was found to be not so low that learning was too slow, while permitting training losses to indeed decrease and learning to occur.

Parameter	Value
Maximum initial learning rate	0.001
Gamma by which learning rates were multiplied after each epoch	0.975
Loss criterion	MSE
Gradient descent : validation split of training simulations	14:3

Table 5: Parameters used to train the network at this training stage with C3,0-uncorrected STEM Ronchigram simulations.

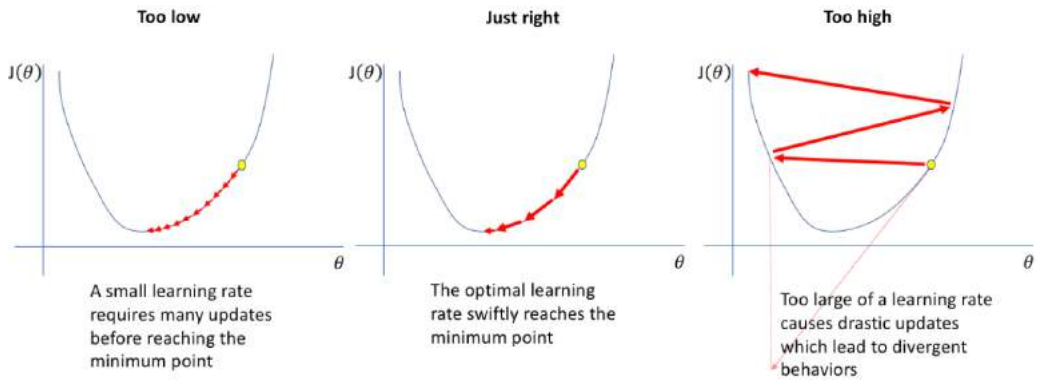


Figure 34: Depiction of how learning rate influences gradient descent. Figure from [55], permission for reuse requested from author Jeremy Jordan.

It was expected that the visibility of C1,2 in the average simulation (which was cropped to highlight this) and the acceptable learning rate would permit extensive training to recognise c1,2 and ϕ1,2 in simulations in Table 4 to occur satisfactorily. Results for this training are as follows, first for 85,000 simulations, a quantity which worked previously.

7.2.1 Training with 85,000 Simulations

In the first training run to evaluate C1,2 in the aforementioned simulations, 85,000 simulations were used. Training began with good fit of overall validation loss to training loss (see Figure 35) for approximately 8 epochs (17,500 iterations). It was expected that, after 8 epochs, any recognition the CNN could do would be better for ϕ1,2 than c1,2, since while projected potential's randomness heavily affected the widths of unidirectional streaks (and thus apparent C1,2 magnitude), its effects on their orientation were limited–compare Figures 36 and 37, between which projected potential varies, streaking is perturbed, but streaks remain horizontal.

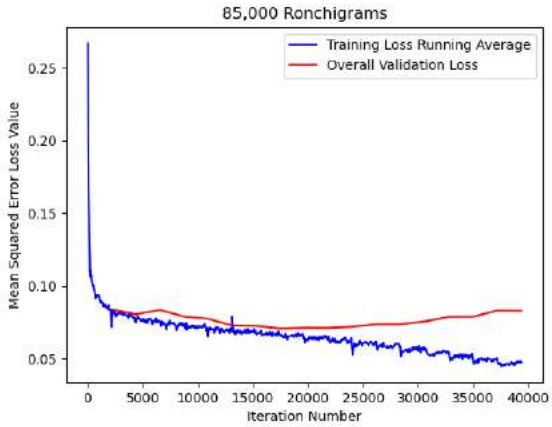


Figure 35: Loss curves for 85,000 simulations. 2,190 iterations occurred per epoch. Note: c1,2 and ϕ1,2 validation loss curves were saved inaccurately due to a bug in the code, so are not shown.

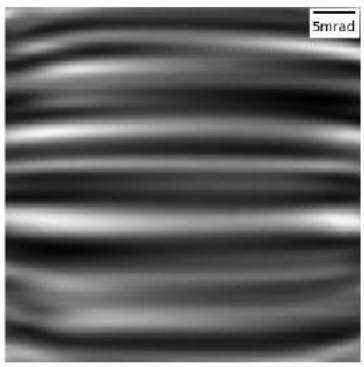


Figure 36: Cropped region of Figure 31 for one projected potential.

To test this, trend graphs (see Section 6.3) were generated for c1,2 and ϕ1,2 using minima and maxima in Table 4, using the version of the CNN, regarding its weights, after

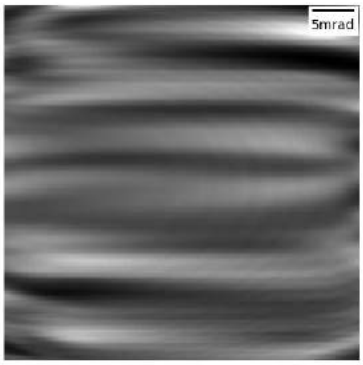


Figure 37: Figure 36 with a different projected potential but other parameters unchanged.

the 8th epoch–see the inflection point of the overall validation loss curve in Figure 35, where generalisation ability is expected to be the best.

The resulting trend graphs in Figures 38 and 39 suggest the CNN was indeed learning ϕ1,2's trends more easily than c1,2's, trends being recognised in Figure 39 former but trend recognition hardly identified in Figure 38 besides a few anomalies approaching the line. Additionally, predictions in Figure 39 seem to band towards the midpoint of the ground truths–retrospectively, this may be due to the ability of the CNN interpolating between its learning of more extreme ϕ1,2 values to assist its learning of medium such values, therefore enhancing its ability to correctly predict ϕ1,2 in simulations where it's value is middling. Regardless, this banding can be seen for later figures too and indicates that randomly sampling ϕ1,2 using a probability distribution biased towards ends of the range may be warranted for more uniform predictions in future work.

However, in Figure 35, overfitting became clear from roughly the 9th epoch (after 19,700 iterations) onwards, where overall validation loss begins to rise. Overfitting for C1,2 overall didn't appear as soon as in Figures 29 and 30, where the deviation's onset occurs after only 5 epochs (11,000 iterations) and 3 epochs (6,600 iterations) respectively. This may be because, despite Ronchigram complexity being greater than for the aforementioned loss curves, the lack of highly-aberrated regions shown to the CNN in the current training run compared to that for Figure 29 makes generalisation easier, while the increased discernibility of aberration magnitude and angle from C1,2's streaking extent and orientation respectively makes learning magnitude and angle easier than for C2,3's pattern in Figure 28, thus leading to better losses than in Figure 30.

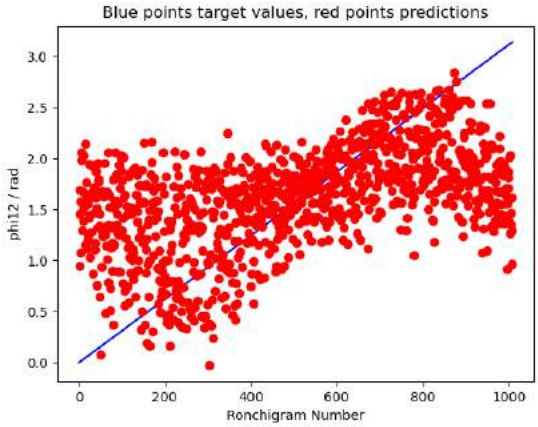


Figure 38: c1,2 trend graph after 8 epochs of training for the network trained using 85,000 C3,0-uncorrected STEM simulations. Note: 1 anomaly is outside of the dimensions of this plot.

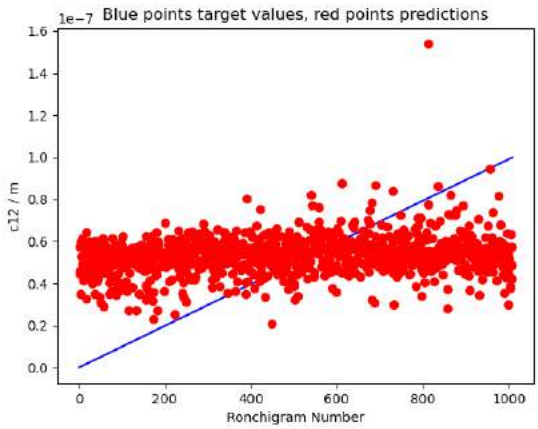


Figure 39: ϕ1,2 trend graph after 8 epochs for the network trained using 85,000 C3,0-uncorrected STEM simulations. Note: 1 anomaly is outside of the dimensions of this plot.

Regardless, efforts were required to prevent the early overfitting in Figure. It was thought the best solution to improve generalisation would be, during training, expose the CNN to more of the aberration value combinations that are possible. This was especially necessary given how large the *input space* was–simulations here feature all 1st-5th order aberrations, including 14 magnitudes and 11 angles. Therefore, 80,000 simulations were to be replaced with 500,000 simulations to better represent the input space and lead to better generalisation, as expected to be indicated by later overfitting (if overfitting was seen) in subsequent training runs. Furthermore, it was also expected that having more validation images would mean validation loss curves depicted CNN generalisation ability more reliably, since validation images would better cover the space of possible simulations. Additionally, c1,2 evaluation was expected to significantly improve, given the exposure of the CNN to more examples of random projected potential for a given similarity of aberration constants and a given c1,2 in the ground truth. Results are in the following section.

7.2.2 Training with 500,000 Ronchigrams

Training was re-initiated sing the new 500,000 simulations, made using the same parameters as in Table 4. raining parameters from Table 5 were reused and consecutive runs were done, each loading the result (in weights, learning rate information and optimiser settings) of its predecessor. The curves in Figure 40 contain stitches from these consecutive runs, individual validation losses for c1,2 and ϕ1,2 displayed too this time.

No overfitting is visible among any of Figure 40's 61 epochs, with all validation loss curves following training loss closely. This beats Figure 35, where overfitting is seen after only around 9 epochs, validating the increase in training data for generalisation improvement. Furthermore, trend checking was repeated after the 8th epoch for comparison with Figures 38 and 39. Figures 41 and 42 depict the result of trend-checking, trends now seemingly noticed to a higher degree than for the previous trend graphs. As can be seen from Figure 41, the did indeed begin to recognise trends in c1,2 as it was hypothesised more training data would permit–therefore, it may have begun to learn to separate projected potential from its search for C1,2's streaking, although at this stage it seems ϕ1,2 trends are still being recognised more easily than c1,2's.

After seeing in real-time the promising trend graphs and good fit of the loss curves,

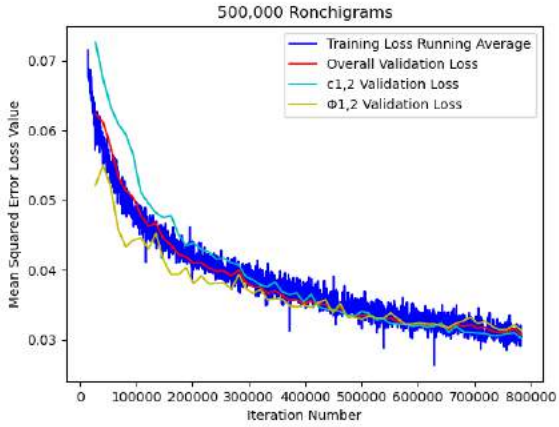


Figure 40: Stitched loss curves for the 500,000-simulation training run. 12,900 iterations per epoch. The curves here begin at the values found at the end of the first epoch, the rest being in Figure 81 in Section 13 because of its unimportance and lack of fit with the area of the curve displayed in the current figure.

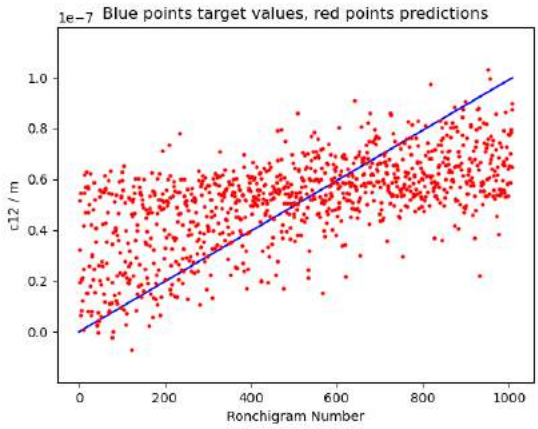


Figure 41: c1,2 trend graph after 8 epochs of training for the network trained using 500,000 C3,0-uncorrected STEM simulations. Note: 1 anomaly is outside of the dimensions of this plot.

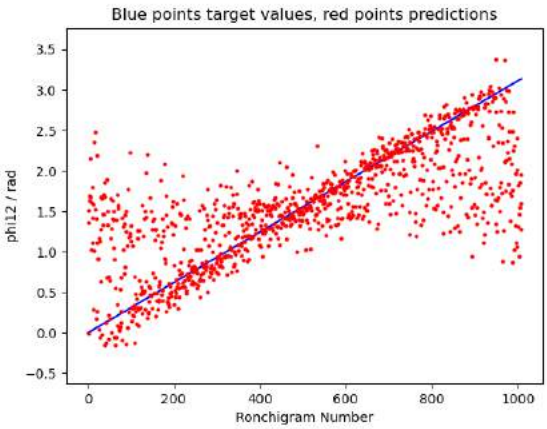


Figure 42: ϕ1,2 trend graph after 8 epochs of training for the network trained using 500,000 C3,0-uncorrected STEM simulations.

training was continued until the end of Figure 40, where trend checking was again done. It was expected that even better fit of predictions to linear ground truths would be achieved by the latest CNN, and this is indeed confirmed by Figures 43 and 44. For Figure 43, while anomalies are still present, predictions are concentrated closer to the c1,2 ground truth line than in Figure 41. Similarly, the same can be said for the comparison of Figure 44 to Figure 39. Therefore, the latest trend graphs demonstrate the success of continued training to recognise c1,2 and ϕ1,2 among the 500,000 simulations used for training.

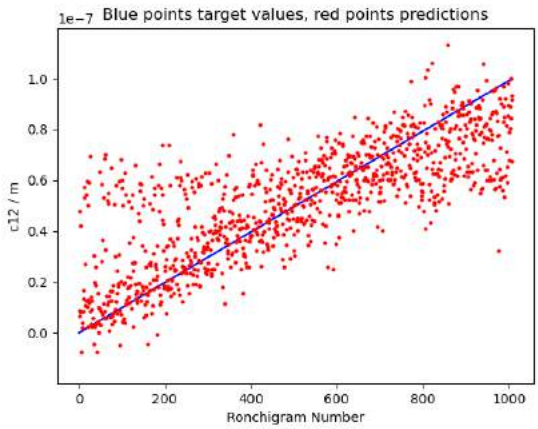


Figure 43: c1,2 trend graphs at latest point in training for the network trained using 500,000 C3,0-uncorrected STEM simulations.

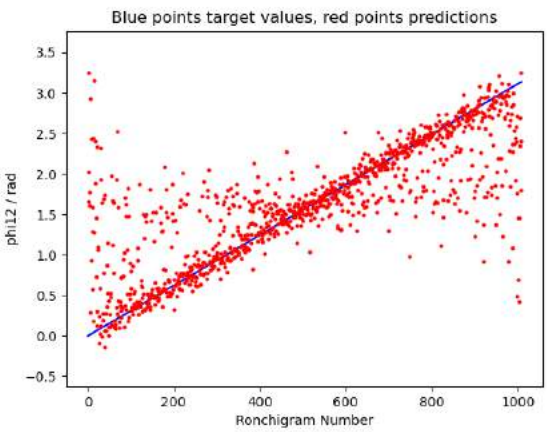


Figure 44: ϕ1,2 trend graph at latest point in training for the network trained using 500,000 C3,0-uncorrected STEM simulations.

Given the success of the CNN in trend-recognition for c1,2 and ϕ1,2 in simulations emulating the C3,0-uncorrected STEM, the closeness of predictions to reality was now visually depicted by generating the comparisons described in Section 6.3. To do this, 100 new simulations (randomly sampling from Table 4 again) were generated and put into

the comparison pipeline, Figures 45 to 48 being four randomly-chosen resulting comparisons of test Ronchigram to predicted Ronchigram. The predicted Ronchigrams roughly approximate ϕ1,2 in most of the four cases, with respective AEs of 0.06π rad, 0.31π rad, 0.05π rad and 0.02π rad, these errors respectively 6%, 31%, 5% and 2% of the width of the range of possible values. The second of these errors corresponds to Figure 46, where the discrepancy between test ϕ1,2 and prediction is glaring in that the streaking is nearly orthogonal between the test and predicted Ronchigram, this discrepancy possibly corresponding to an anomaly similar to those in Figure 44.

In comparison, it is difficult to judge prediction success for c1,2 visually due to projected potential differences admitted in Section 6.3. In Figures 45 to 48, predicted c1,2 differs from ground truth by 25.4%, 28.6%, 38.6% and 3.2% respectively. While the former three values warrant further training, the fourth (coupled with the corresponding aforementioned 2% error for ϕ1,2) at least demonstrates the CNN's capability to evaluate c1,2 and ϕ1,2 highly-accurately in the same Ronchigram simulation here, which comparison inference was established to do. Thus, at this stage of incremental development towards prioritising success in the C3,0-corrected STEM instead,it was decided that further work on the C3,0-uncorrected STEM would be halted in the interest of time.

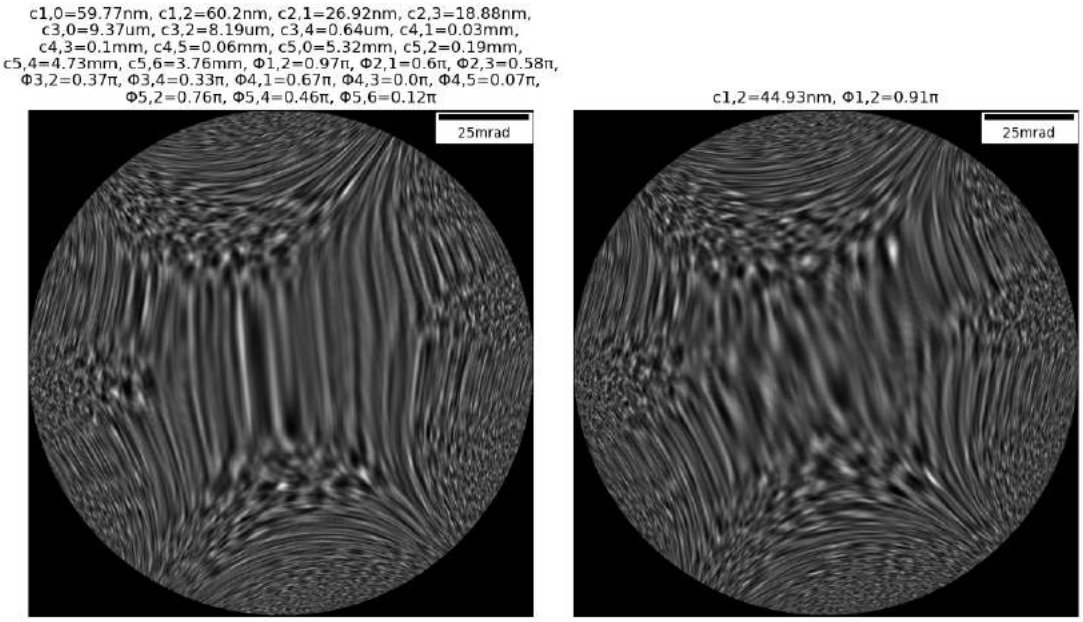


Figure 45: A comparison of test simulation (left) with predicted simulation (right), generated as in Section 6.3, for the latest point in training on the 500,000 simulations emulating the C3,0-uncorrected STEM.

However, before discussing results of training using simulations emulating the C3,0-

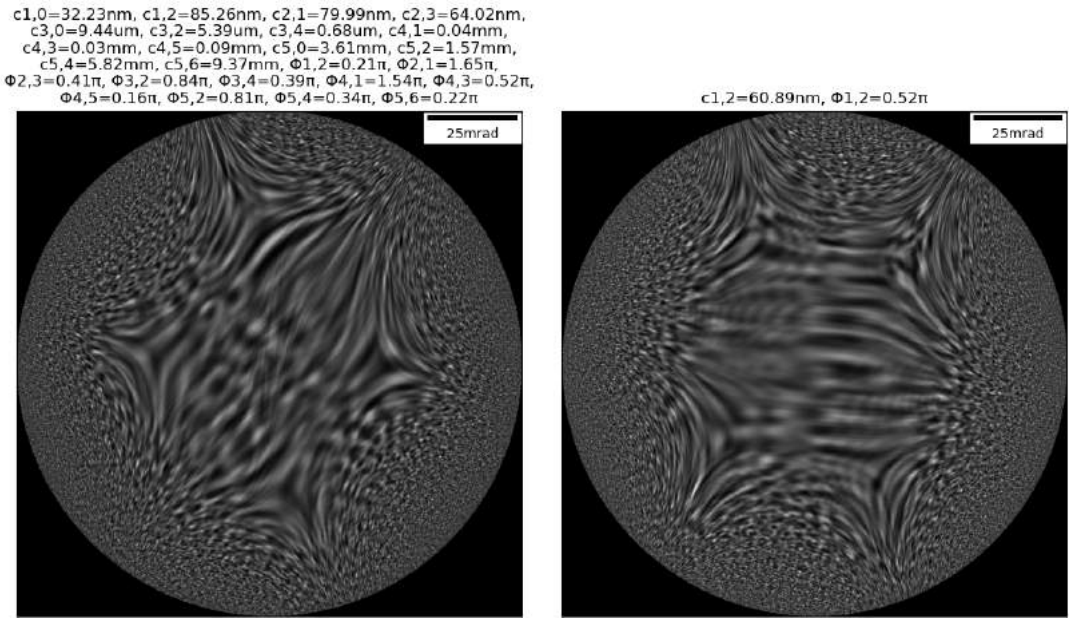


Figure 46: A comparison of test simulation (left) with predicted simulation (right) for the latest point in training on the 500,000 simulations emulating the C3,0-uncorrected STEM.

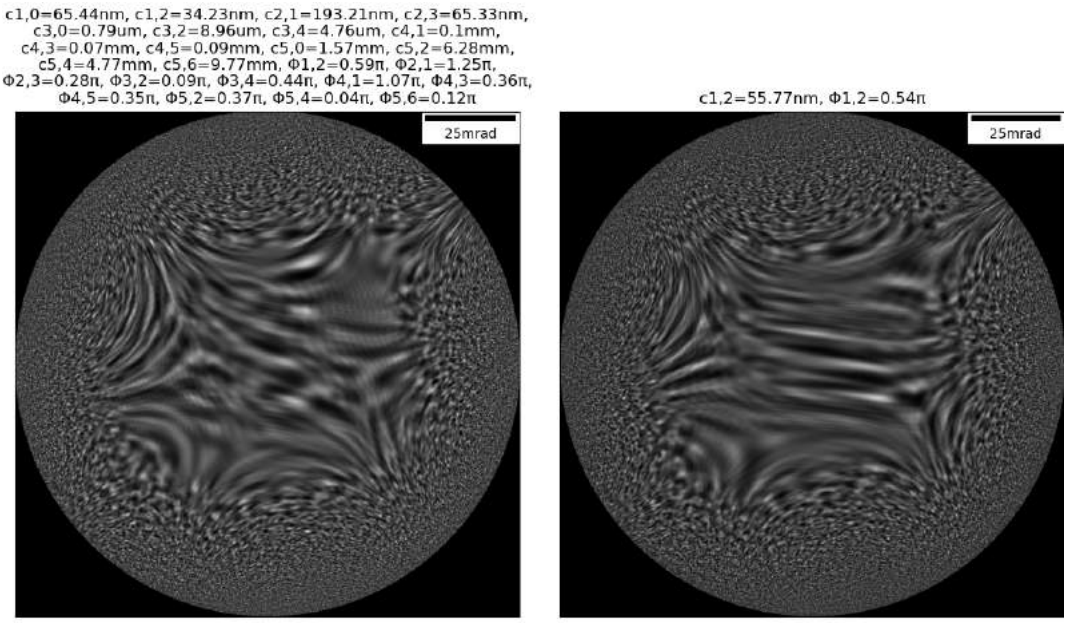


Figure 47: A comparison of test simulation (left) with predicted simulation (right) for the latest point in training on the 500,000 simulations emulating the C3,0-uncorrected STEM.

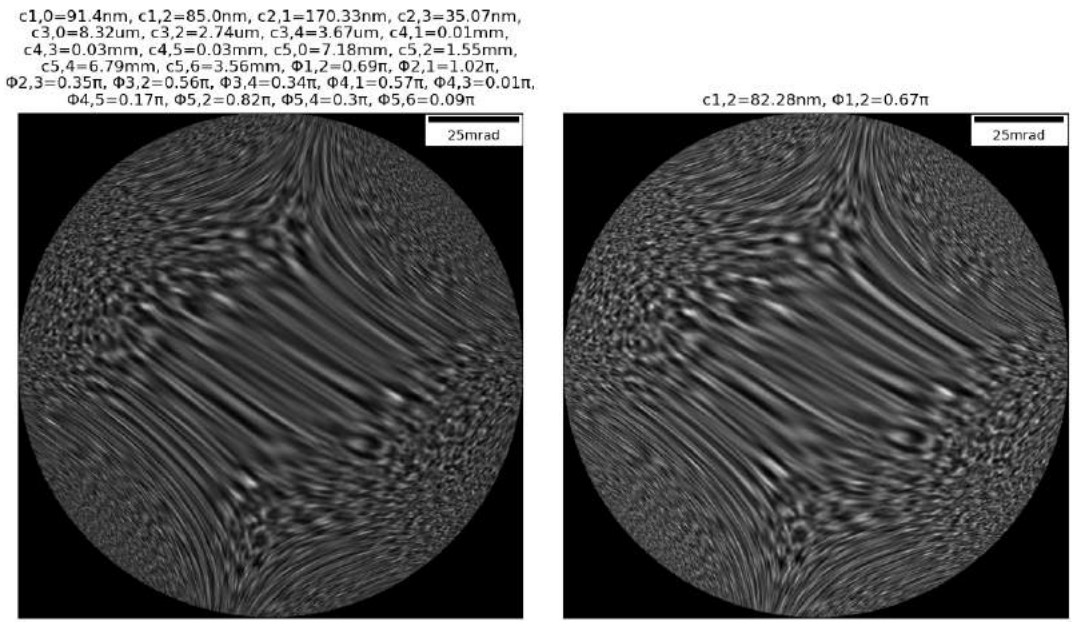


Figure 48: A comparison of test simulation (left) with predicted simulation (right) for the latest point in training on the 500,000 simulations emulating the C3,0-uncorrected STEM.

corrected STEM, it was desired to investigate the depiction in Figure 40 of validation loss eventually becoming lower c1,2 than for ϕ1,2, from 51 epochs (658,000 iterations) onwards. Because this happened late in this training run, it was hypothesised that this might be due to the following notion:

1. Over time, the CNN continues to learn to recognise how high c1,2 is (and thus c1,2 validation loss continues to decrease) by analysing the extent of streaking seen; it learns that a lot of streaking suggests high c1,2, whereas little streaking suggests low c1,2.

2. Simultaneously, the CNN continues to learn to recognise ϕ1,2 (and thus ϕ1,2 validation loss continues to decrease) by analysing the orientation of streaking seen; however, where c1,2 is low, streaks become discernible and thus ϕ1,2 becomes difficult to learn and predict.

It was thought that, while supposition *2.* might not cause learning of ϕ1,2 to predominate early in training due to projected potential issues the CNN must first navigate to learn c1,2 (see Section 7.2.1), the CNN could still separate random variations in projected potential from streaks with sufficient training, as implied by trend-checking improvements from Figure 38 to Figure 41. On the other hand, regarding recognising ϕ1,2 when c1,2 was

low, the CNN was not expected to as easily notice what was barely present at all, thus meaning c1,2 validation losses would eventually become lower than ϕ1,2's late in training.

To test this rationale and to decide if the sampling of c1,2 should have been skewed to lower values to better train the network to recognise ϕ1,2 on the lower end of the spectrum of C1,2 magnitudes, a bar plot was constructed by creating 10 sets of 1,000 simulations. In each, random sampling was done of all simulation parameters from ranges in Table 4, except the range for c1,2 was split into ten sub-ranges, each for a different simulation set. This would allow the predictions of c1,2 and ϕ1,2 across simulations with varying magnitudes of C1,2 to be compared, by allowing the latest stage of the C3,0-corrected-STEM CNN to perform inference on each set and provide a success metric for each, demonstrated by Figure 49.

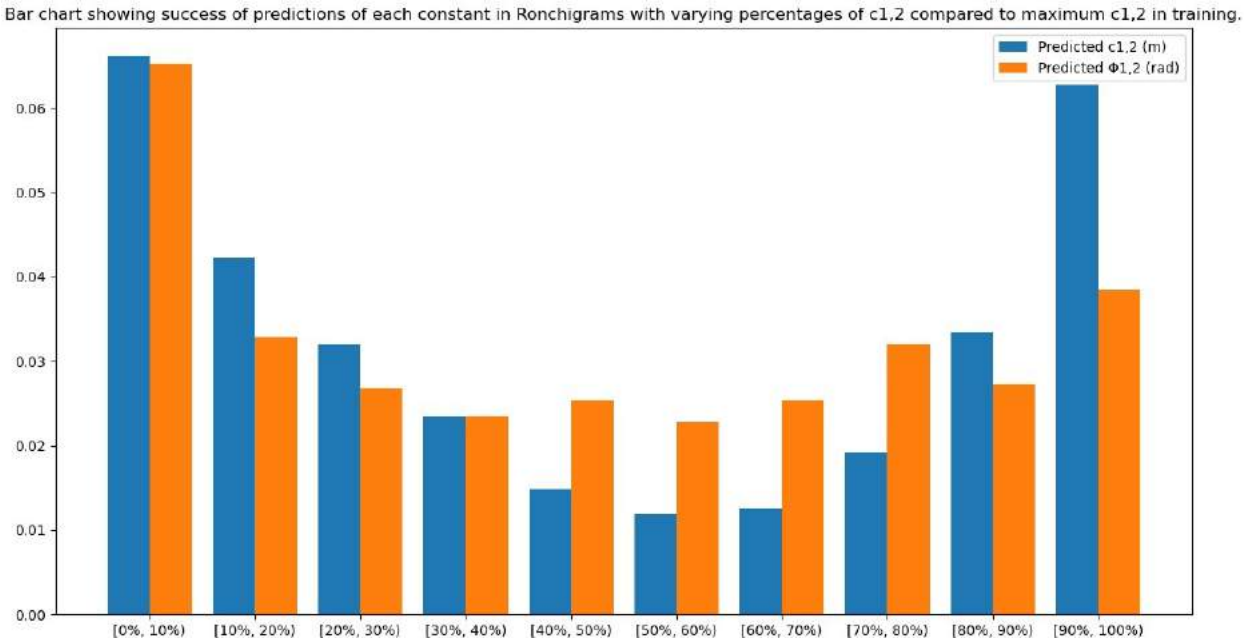


Figure 49: Bar plot showing the success of predictions of c1,2 and ϕ1,2 in Ronchigrams of varying c1,2. Percentages here are those of the maximum value in training (100nm) that c1,2 has in inference data. To generate each bar, inference was done on 1,000 Ronchigrams with c1,2 randomly sampled for the corresponding percentile range and all other factors randomly sampled from ranges in Table 4. Vertical axis is the MSD between predicted values and ground truths.

Contrary to hypothesis, the bar plot doesn't support the idea that the network should struggle to recognise ϕ1,2 over c1,2 overall in simulations featuring relatively small levels of c1,2 because. For sets of test simulation in which c1,2 was between 0% and 30% of its maximum, the overall MSD between values and ground truths was lower for each percentile range when predicting ϕ1,2 compared to predicting c1,2. This suggests continuing to

randomly sample c1,2 using a uniform probability distribution was justified instead of skewing sampling to lower c1,2. However, the dip in prediction error towards the middle of the bar plot for both c1,2 and ϕ1,2 is reminiscent of the banding in Figure 39, and suggests future work might benefit from randomly sampling c1,2 using a probability distribution skewed towards higher and lower bounds instead.

7.3 Training to Evaluate C1,2 in Ronchigrams Emulating the C3,0-Corrected STEM

7.3.1 Training the Network Using Simulations

After the success of training runs to evaluate c1,2 and ϕ1,2 in 500,000 Ronchigrams emulating the C3,0-uncorrected STEM, progression was made towards work on simulations emulating the C3,0-corrected STEM instead but in a quantity of 500,000 once again. These simulations were generated using ranges found in Table 6, differing only from Table 4 via the ranges from which aberration magnitudes were sampled and the detector semi-angle that was simulated. The magnitude ranges were chosen such that the maxima matched those in the set of Ronchigrams experimentally acquired by Dr. Huang, details of this acquisition mentioned in Section 6.5, where aberration values were estimated by JEOL COSMO™. The images encapsulated the desired amount of Ronchigram, quite fully containing the central regions like those seen in average simulations seen earlier in the current Section 7–see Figure 50 for an example.

Since fine-tuning was to be done here, besides matching simulated aberration values and magnitudes to the experimental Ronchigrams, their detector semi-angle was also estimated. This was done by generating a simulation for each Ronchigram based on their quoted parameters then tuning the most accurate simulation to a scale similar to its corresponding experimental Ronchigram by varying its detector semi-angle; the detector semi-angle achieving the closest comparison was 180mrad, as seen in Figures 50 and 51.

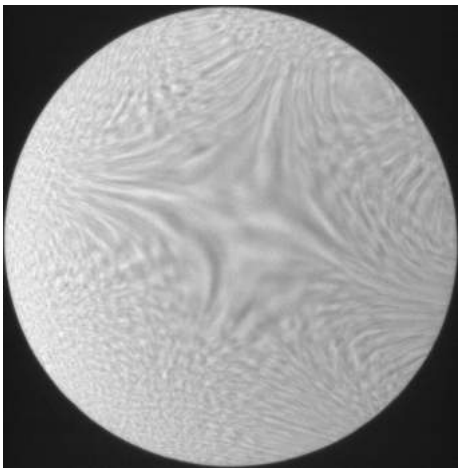


Figure 50: Experimental Ronchigram 10 (see Table 7), scale bar omitted due to lack of knowledge of simulation dimensions before comparison with Figure 51. Brightness here is enhanced by ImageJ here relative to illumination seen in practice.

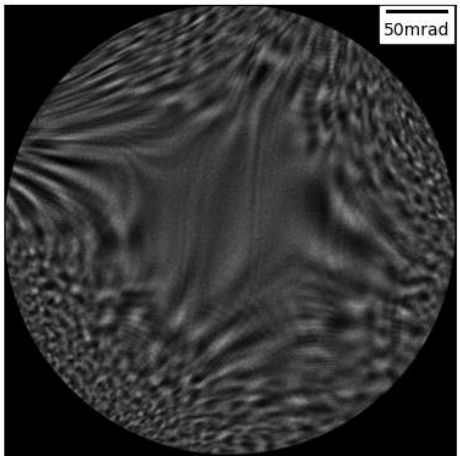


Figure 51: Simulation made using parameters returned by JEOL COSMO™ for Ronchigram 10 (depicted in Figure 50).

Parameter	Value or range
I range	[100pA, 1nA)
t range	[0.1s, 1s)
c1,0 range	[0, 7.619nm)
c1,2 range	[0, 2.449nm)
c2,1 range	[0, 33.883nm)
c2,3 range	[0, 13.849nm)
c3,0 range	[0, 1.045um)
c3,2 range	[0, 0.277um)
c3,4 range	[0, 0.28um)
c4,1 range	[0, 4.159um)
c4,3 range	[0, 1.951um)
c4,5 range	[0, 0.791um)
c5,0 range	[0, 0.047mm)
c5,2 range	[0, 0.000mm)
c5,4 range	[0, 0.000mm)
c5,6 range	[0, 0.010mm)
$\phi_{n,m}$ range	[0, $\frac{2\pi}{m}$rad)
Detector semi-angle	180mrad

Table 6: Parameters used to simulate the C3,0-corrected STEM Ronchigrams used to train the network at this stage. In the values of aberration angles here, m refers to the order of rotational symmetry.

After generating simulations, attempts were again made to select the area of the simulation shown to the network during training by perturbing c1,2 and ϕ1,2, by 50% in each case, from their values in the average simulation (shown in Figure 52) and seeing which area of it was affected in each case. The blue box drawn over the average simulation and in Figures 53 and 54 shows the region thereby selected, with the increased c1,2 hardly informing this selection based on its negligible effect of perturbing it from the average simulation here.

Next, training was re-initiated using parameters in Table 5, except the most recent set of weights (without the learning rate and optimiser information saved alongside) corresponding to Figure 40 were first loaded before training began. This was because, while dimensions were different for the C3,0-corrected STEM simulations used than for the C3,0-uncorrected STEM simulations, the aberration value ranges in Table 6 are but subranges of those in Table 4, meaning the aforementioned weights should already have been trained to some extent to have capabilities in evaluating c1,2 and ϕ1,2 among aberration combinations in the new simulations. Given the increased number of images occupying the ranges here compared the same ranges in Section 7.2.2, it was expected that further training would only develop these capabilities, and achieve superior training results compared to those in Figure 40 due to the decreased input space.

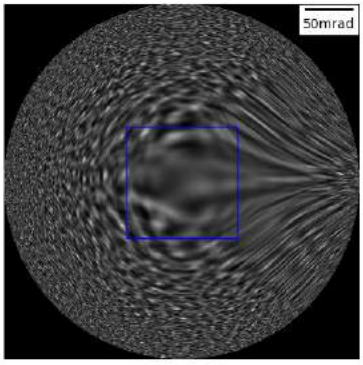


Figure 52: The average simulation from ranges in Table 6, the crop being the region shown to the CNN in training. The crop emulates an aperture-cropped simulation with an 80mrad detector semi-angle.

Training was done for 28 epochs only due to time constraints, loss curves associated with this training run can be seen in Figure 55. Although all curves trend downwards overall in MSE, the CNN begins to overfit regarding overall validation loss after around

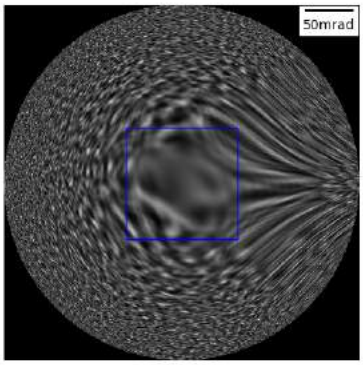


Figure 53: Figure 52 with ϕ1,2 increased by 50%.

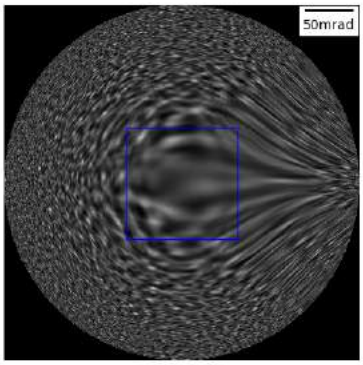


Figure 54: Figure 52 with c1,2 increased by 50%.

16 epochs (206,400). However, overfitting in ϕ1,2 occurs sooner, clear divergence seen between the ϕ1,2 validation and training loss curves seen after only 4 epochs. In contrast, overfitting isn't seen for c1,2, since even at the end of training the c1,2 validation loss curve is slightly below the training loss running average; this might have been because, while c1,2 validation loss isn't averaged over ϕ1,2, whose validation loss is comparably high, training loss running average is indeed averaged over training losses for ϕ1,2, which the CNN may be struggling to learn to evaluate. Because learning to recognise c1,2 seemed to be going so well, training was continued beyond the onset of overall overfitting.

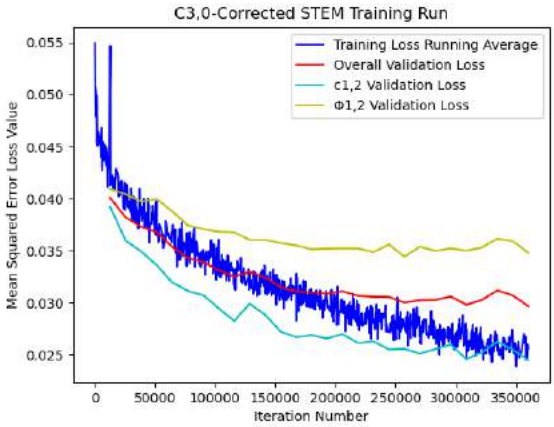


Figure 55: Loss curves overlapped for the CNN being trained with simulations emulating the C3,0-corrected STEM. 12,900 iterations per epoch.

At the end of this training run, it was anticipated that the significant lowness of MSE validation loss for c1,2 compared to that for ϕ1,2 meant the CNN would more accurately identify trends in test data for the former than for the latter–after all, said losses were calculated via inference on data not shown to the network during gradient descent. In order to investigate this, trend graphs were formed using the weights at the latest point in the training run depicted by Figure 55, results of which can be seen in Figures 56 and 57. In both cases, the network seems to be approaching the ability to predict trends correctly overall, however, contrary to hypothesis, it is not so conclusive that the network performs better at recognising trends in c1,2 than in ϕ1,2, since there seem to be a greater density of prediction points closer to the line of ground truths in Figure 57 than in Figure 56. However, while this is the case, the shape of the predictions graph for c1,2 seems more uniform than that for ϕ1,2, where anomalies deviate from the linear trend relatively significantly, especially near the upper and lower bounds of ϕ1,2.

It could be that such values that are significantly hampering the decrease in validation loss for ϕ1,2 shown in Figure 55, especially given how punitive the MSE loss metric is of predictions far from actual values due to the squaring that it does. However, large prediction errors may be more crucial to be rid of than smaller ones, especially since they may be more problematic when shifting aberration values in quick, repetitive corrections. Therefore, the aforementioned punitive nature of MSE loss suggests it may be a criterion worth continuing with in future work, since it may eventually ensure large discrepancies are avoided at all by ensuring the true loss minima in gradient descent correspond to weight predicting a greater proportion of acceptably accurate aberration values.

Figures 58 and 59 are trend graphs made exactly the way Figures 56 and 57 were, except with simulations done to a detector semi-angle 80mrad instead and then cropped to the inner square of the condenser aperture. These confirm the supposition in Section 6.1.3 that the final network trained for the C3,0-corrected STEM should work on the aforementioned simulations, in which the average probe is acceptable and doesn't depict oversampling (see Figure 22). Predictions are generally further from the ground truths for the acceptable-probe pair of trend graphs than for the previous pair, but trends are still starting to be recognised, although further training on the acceptable-probe simulations may be warranted. However, these results demonstrate the network's capabilities on simulations associated with a probe that doesn't feature oversampling, and potential capabilities regardless of simulation dimension as long as Ronchigrams are cropped to an extent consistent with training.

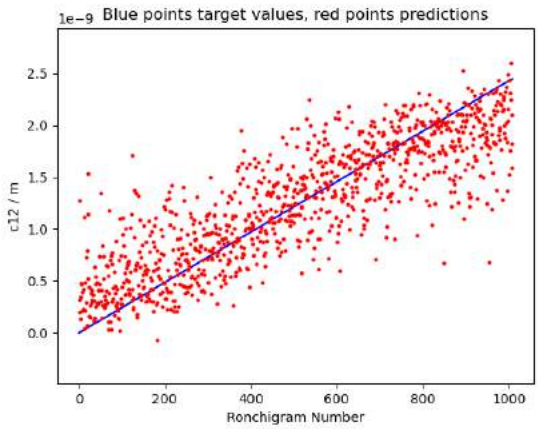


Figure 56: c1,2 trend graph for the latest network trained for the C3,0-corrected STEM, generated using the same type of simulations as for training.

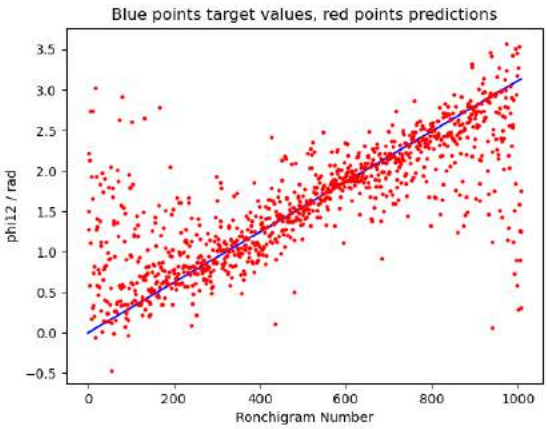


Figure 57: ϕ1,2 trend graph for the latest network trained for the C3,0-corrected STEM, generated using the same type of simulations as for training.

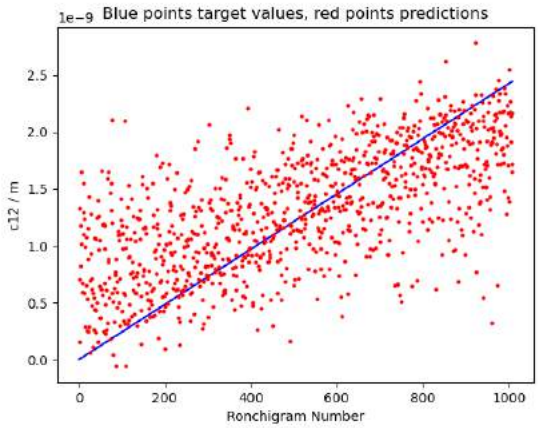


Figure 58: Figure 56 but instead using Ronchigrams simulated to a detector semi-angle of 80mrad and aperture-cropped.

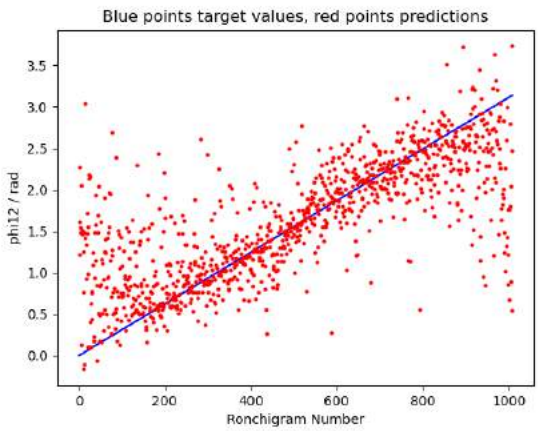


Figure 59: Figure 56 but instead using Ronchigrams simulated to a detector semi-angle of 80mrad and aperture-cropped.

Comparison inference was done on the 180mrad Ronchigrams with the same network used to produce the aforementioned trend graphs, four randomly-chosen examples of which are shown in Figures 60 to 63. The network at least succeeds in simultaneously inferring both c1,2 and ϕ1,2 to high accuracy in the case of Figure 60, with a prediction difference of 6.82% from the ground truth for c1,2 and, for ϕ1,2, an AE only 9% of the range of possible values. High accuracy is not at all the case in Figure 63, where the aforementioned percentages are instead 2800% and 59% respectively. However, in this figure, c1,2 is well below the 1nm value needed for optimal resolution in the STEM[12], so this figure doesn't negate the CNN's applicability in evaluating C1,2 acceptably for aberration correction in practice.

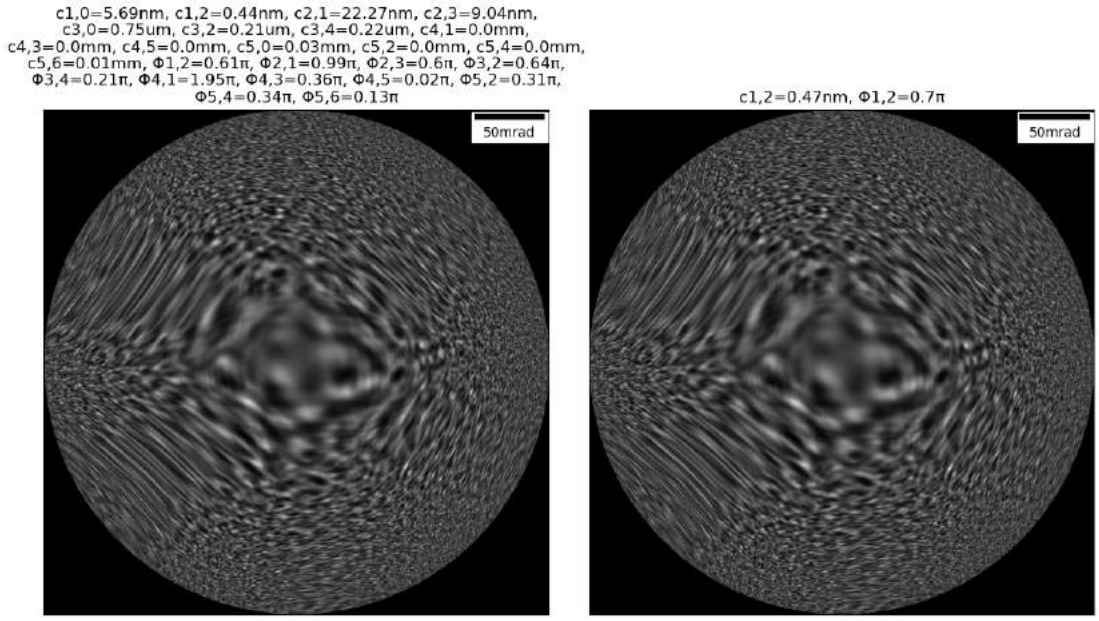


Figure 60: Test Ronchigram (left) vs. predicted Ronchigram (right) generated by the latest C3,0-corrected STEM network using simulations.

While decreasing c1,2 validation losses in Figure 55 would have permitted further learning for c1,2, this was not done due to aforementioned time constraints. At this stage, the network achieves the 15.7% and 18.9% percentage errors mentioned for c1,2 and ϕ1,2 in the Abstract, which may be acceptable where experimental Ronchigrams match simulations and a quick, repetitive correction is possible, as discussed in Section 5. It was decided to investigate how well this success translated to experimental Ronchigrams acquired by Dr. Huang.

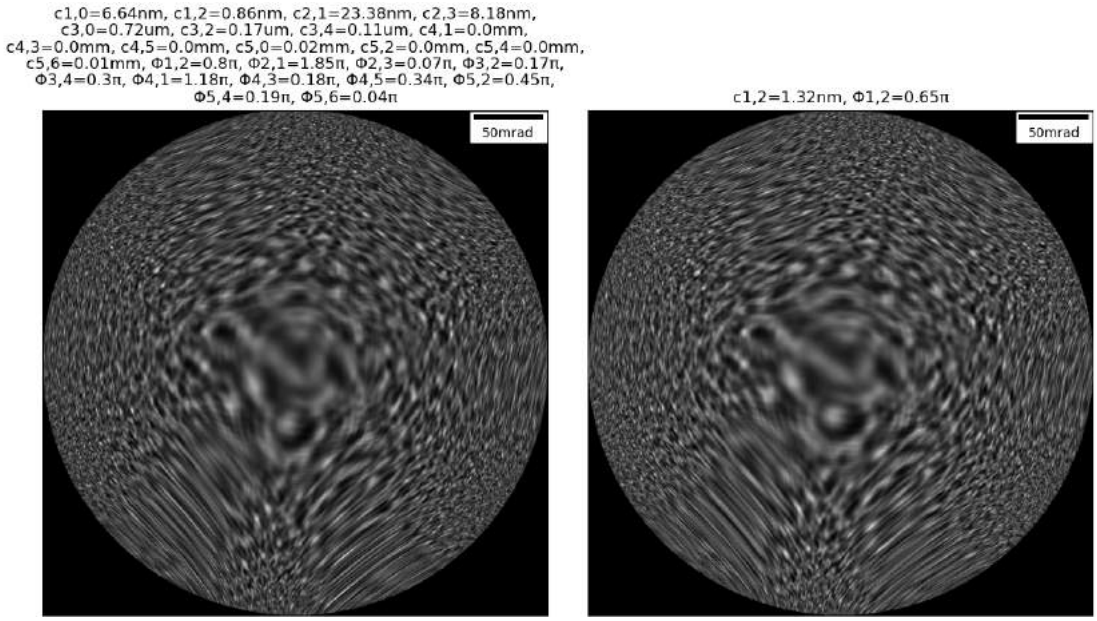


Figure 61: Test Ronchigram (left) vs. predicted Ronchigram (right) generated by the latest C3,0-corrected STEM network using simulations.

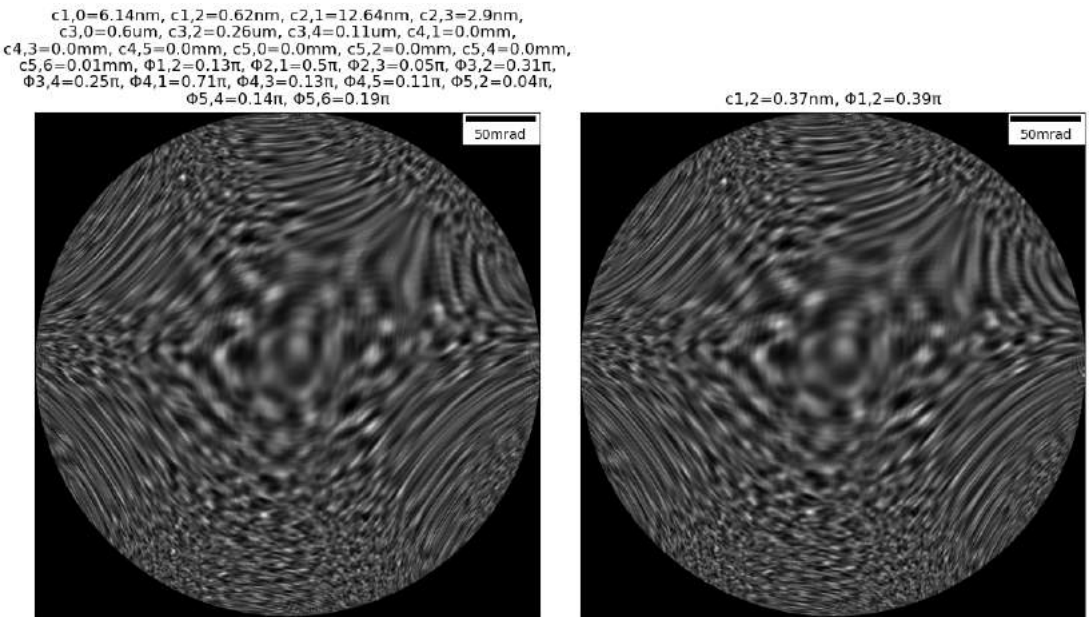


Figure 62: Test Ronchigram (left) vs. predicted Ronchigram (right) generated by the latest C3,0-corrected STEM network using simulations.

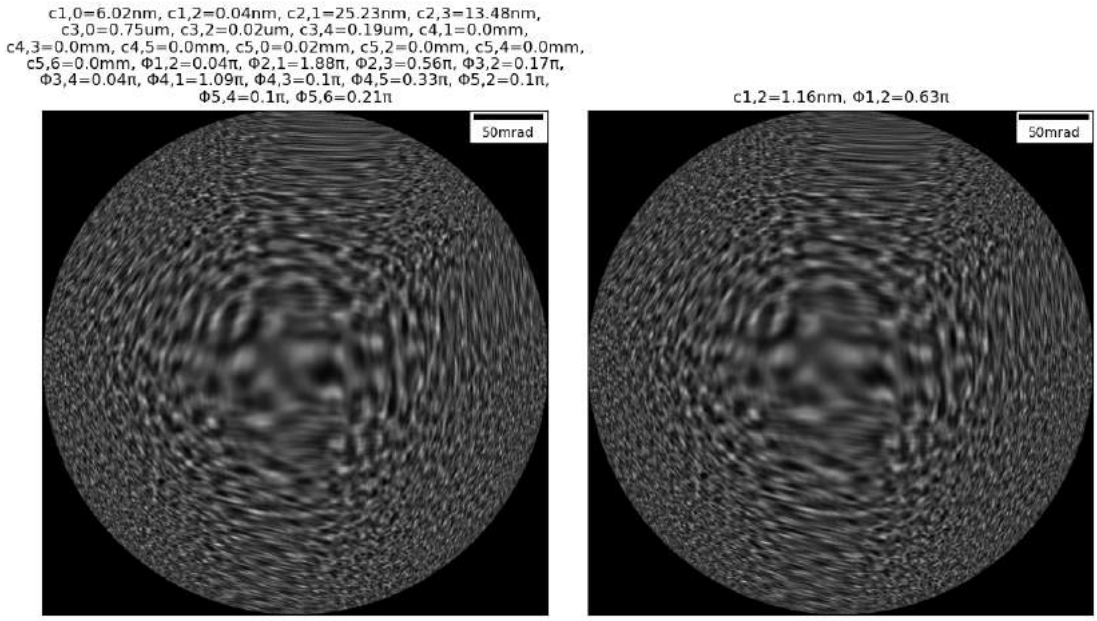


Figure 63: Test Ronchigram (left) vs. predicted Ronchigram (right) generated by latest C3,0-corrected STEM network using simulations.

7.3.2 Experimental Ronchigrams

As mentioned in Section 6.5, there were numerous issues from experimental Ronchigram acquisition, including missing Ronchigram dimensions and electron currents in Amperes, as well as an acquisition specimen not closely matching simulation parameters like expectation value of random phase in projected potential (see Section 6.1.1). Therefore, it was anticipated that the CNN that had succeeded with test simulations in Section 7.3.1 would perform poorly on experimental Ronchigrams. However, if the recognition of any trend was expected, it was the increasing inadequacy of inference as quoted Ronchigram acquisition current percentages (described in Section 6.5) and times decreased, due to the increased corresponding detector noise. For the 11 Ronchigrams acquired, current percentages and acquisition times are listed in Table 7. The aforementioned network was used to generate comparison Figures 64 to 74; maximum electron current assumed to be 1nA (a typical maximum in experimentation, see Section 6.1.2) and the electron percentage in each experimental Ronchigram was applied to this value in generating each predicted simulation. Additionally, predicted c1,2 and ϕ1,2 were plotted alongside estimated for the experimental Ronchigrams by JEOL COSMO™, results of which can be seen in Figures 75 and 76.

Figures 64 to 74 demonstrate the difficulty the network has in recognising C1,2 coupled

Ronchigram Number	Electron Current/%	Acquisition Time/s
1	100	1
2	50	1
3	20	1
4	5	0.1
5	5	0.2
6	1	0.2
7	1	0.2
8	1	0.2
9	1	0.2
10	1	0.2
11	10	0.2

Table 7: Recorded acquisition current percentage and time for each experimental Ronchigram..

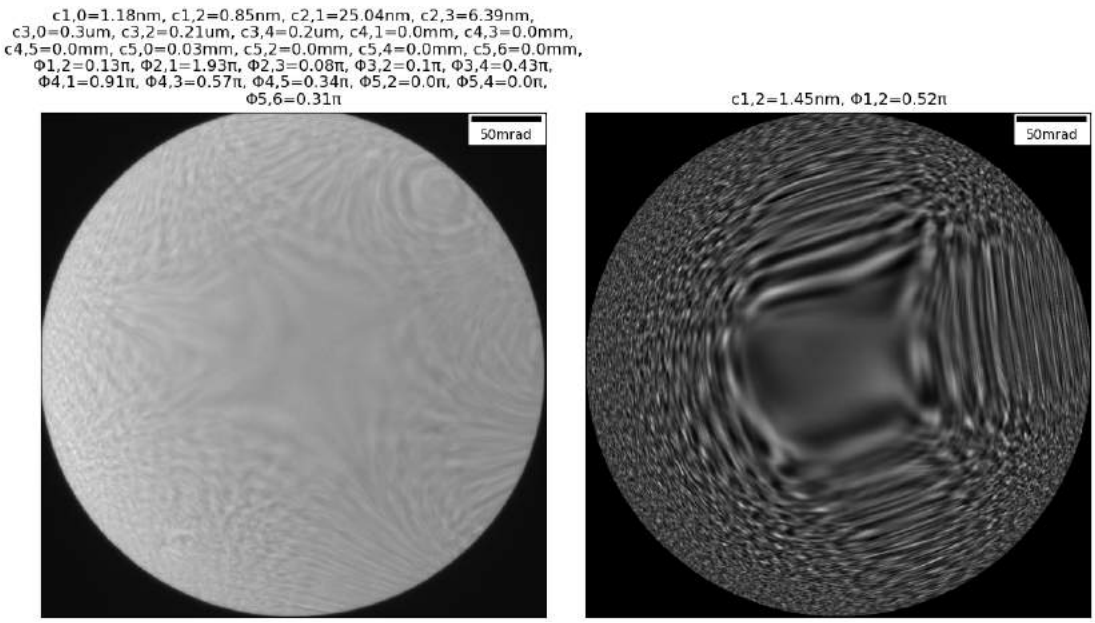


Figure 64: Experimental Ronchigram 1 (left) vs. predicted Ronchigram (right) generated by latest C3,0-corrected STEM network.

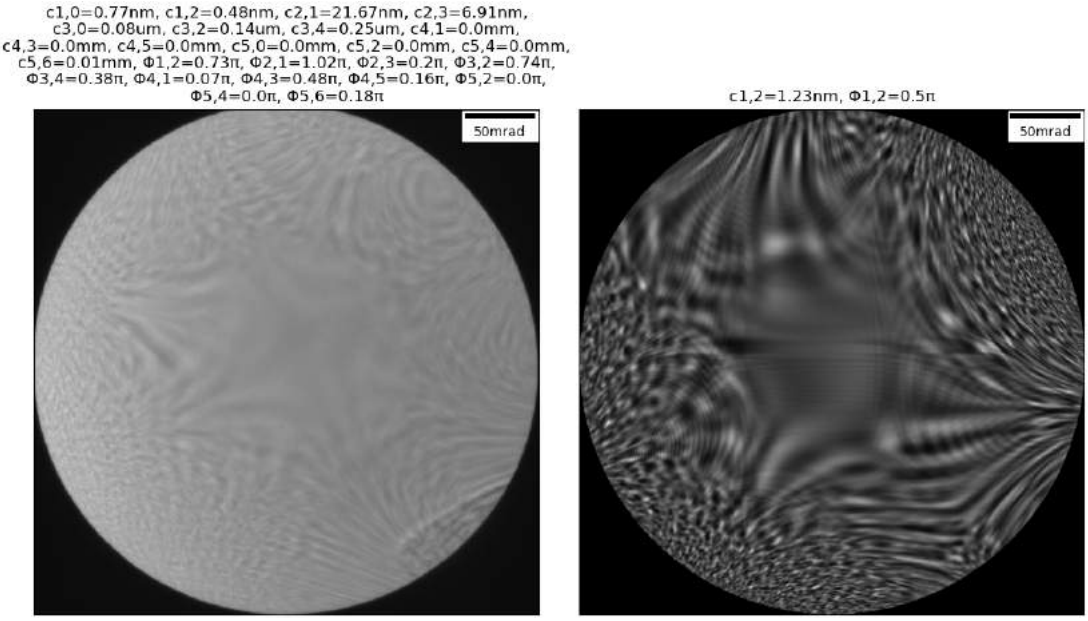


Figure 65: Experimental Ronchigram 2 (left) vs. predicted Ronchigram (right) generated by latest C3,0-corrected STEM network.

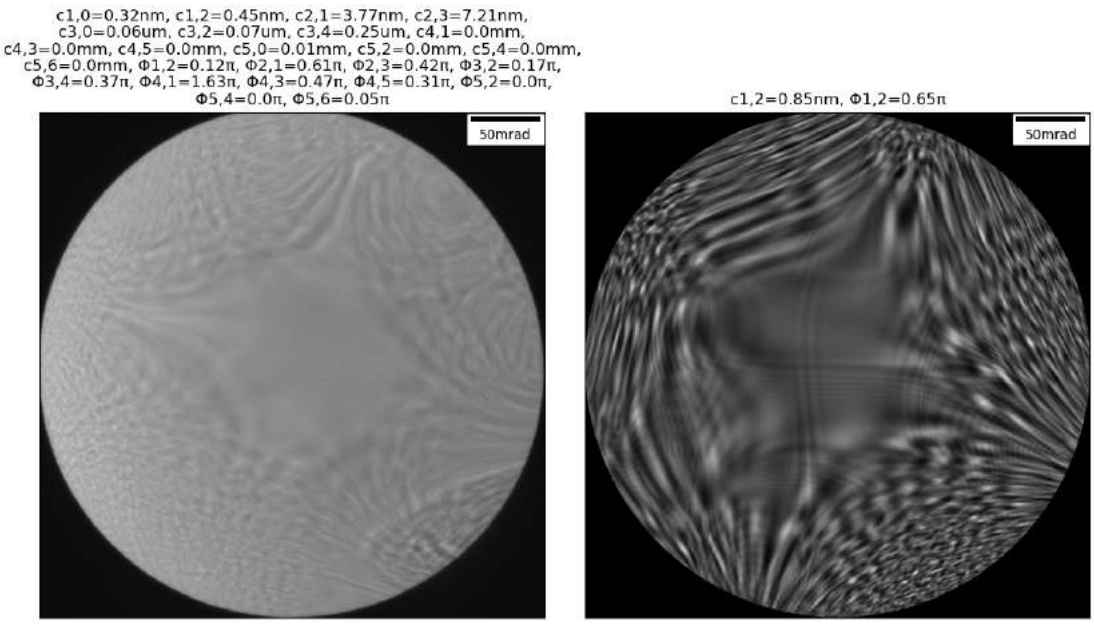


Figure 66: Experimental Ronchigram 3 (left) vs. predicted Ronchigram (right) generated by latest C3,0-corrected STEM network.

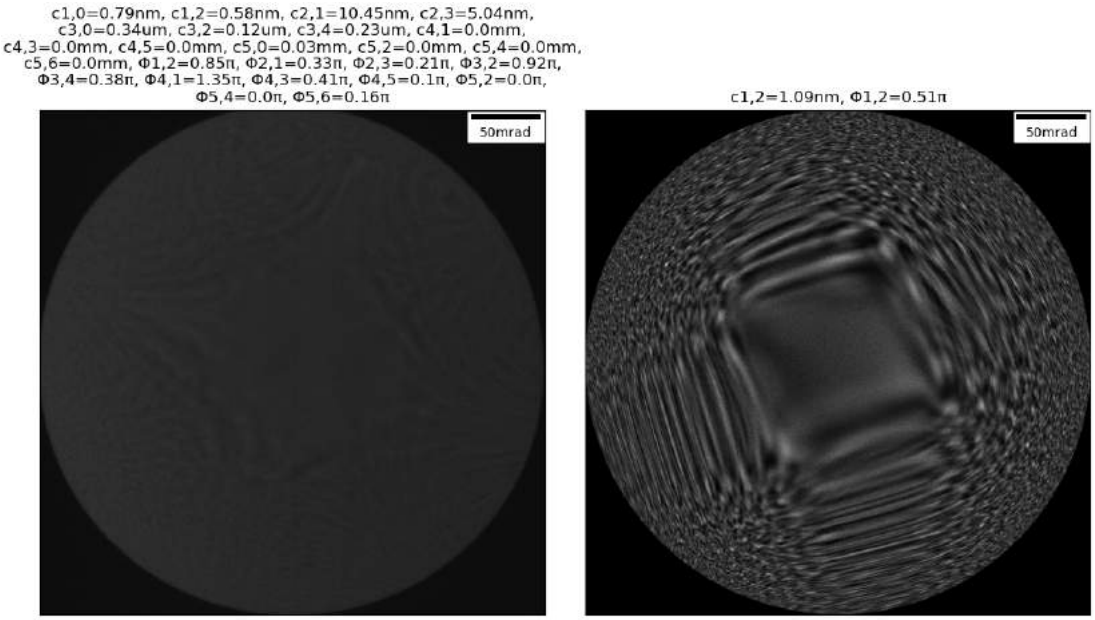


Figure 67: Experimental Ronchigram 4 (left) vs. predicted Ronchigram (right) generated by latest C3,0-corrected STEM network.

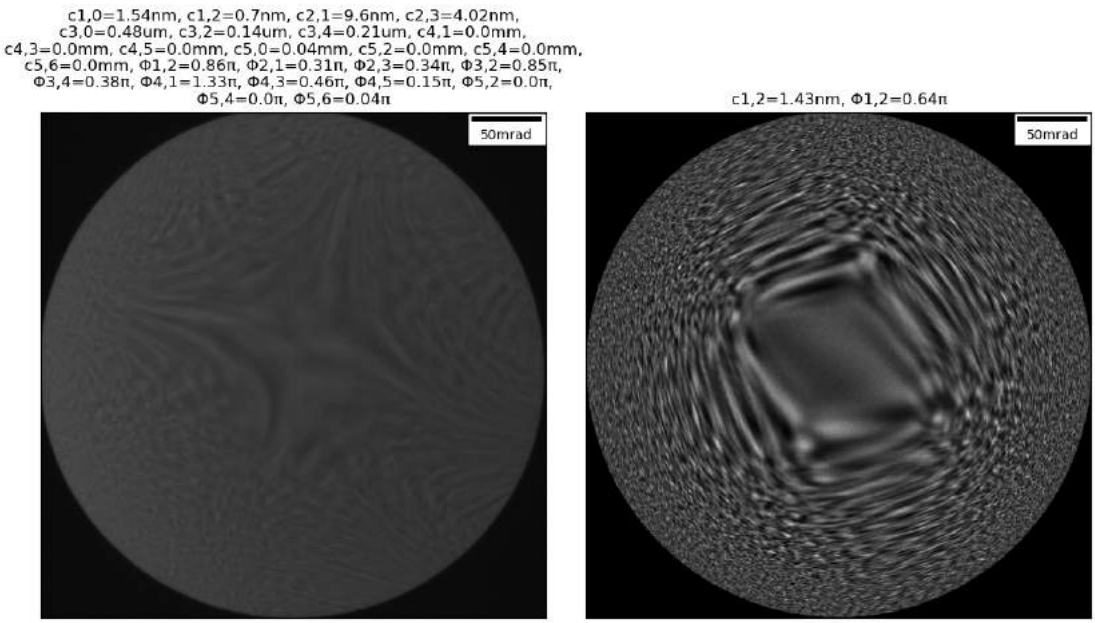


Figure 68: Experimental Ronchigram 5 (left) vs. predicted Ronchigram (right) generated by latest C3,0-corrected STEM network.

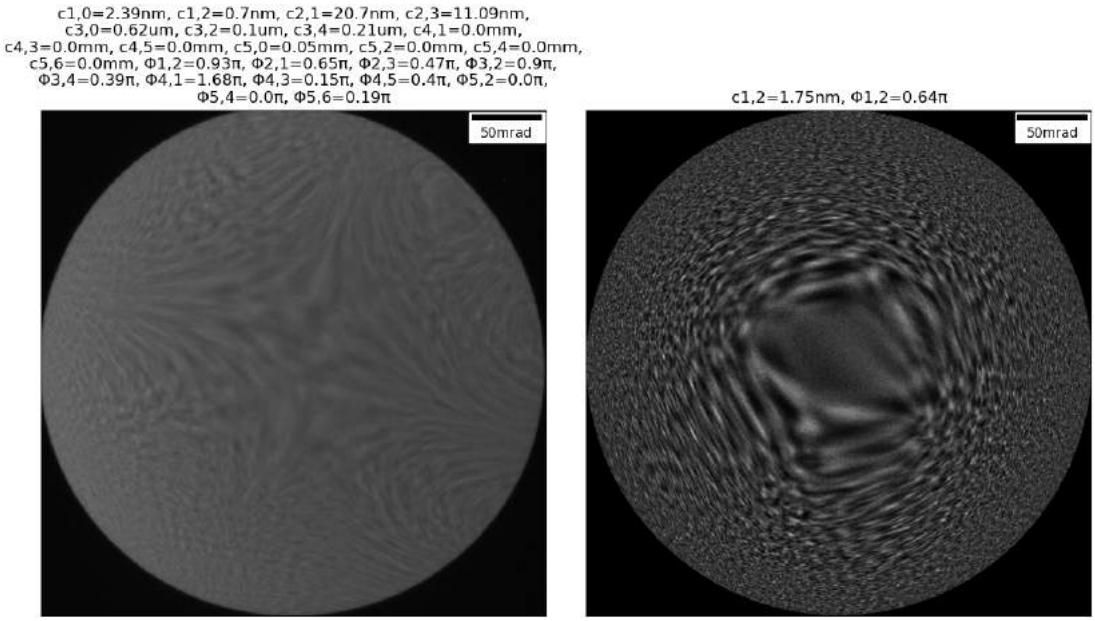


Figure 69: Experimental Ronchigram 6 (left) vs. predicted Ronchigram (right) generated by latest C3,0-corrected STEM network.

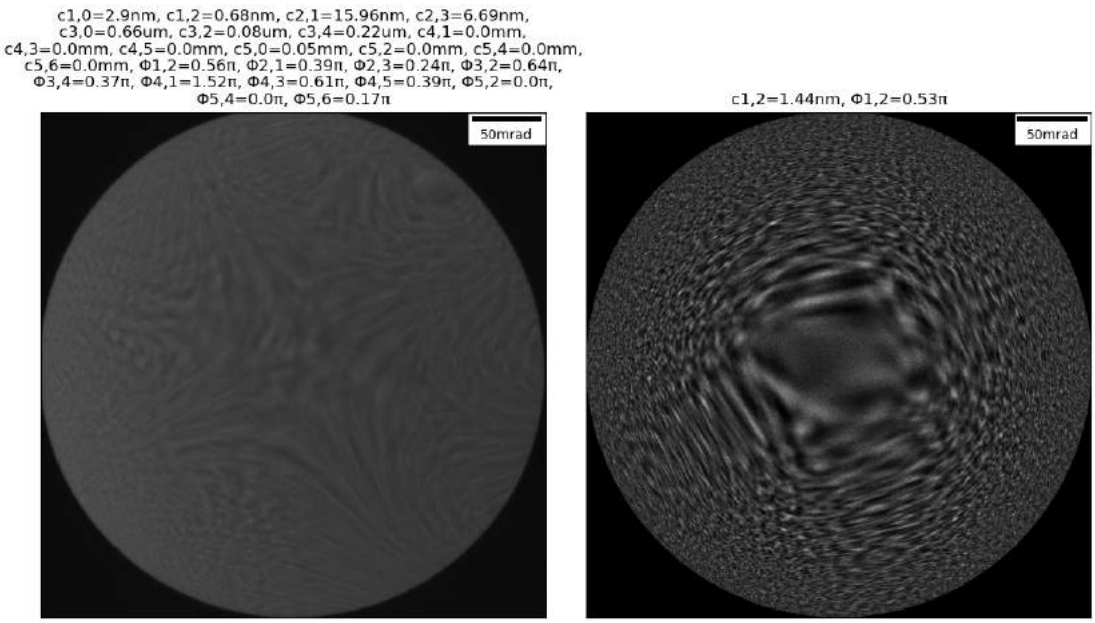


Figure 70: Experimental Ronchigram 7 (left) vs. predicted Ronchigram (right) generated by latest C3,0-corrected STEM network.

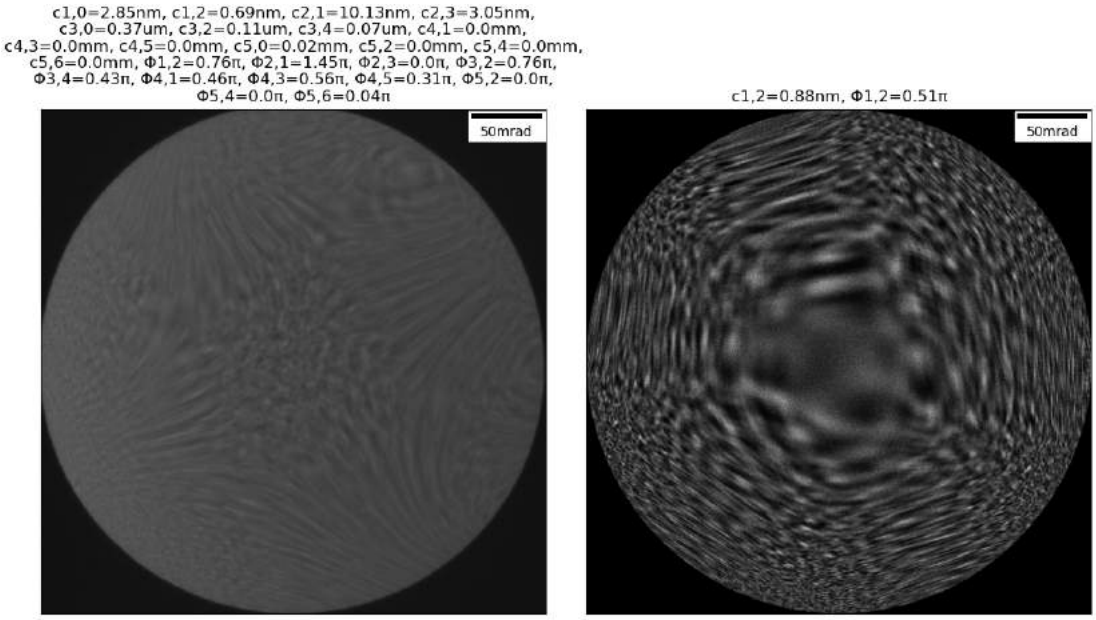


Figure 71: Experimental Ronchigram 8 (left) vs. predicted Ronchigram (right) generated by latest C3,0-corrected STEM network.

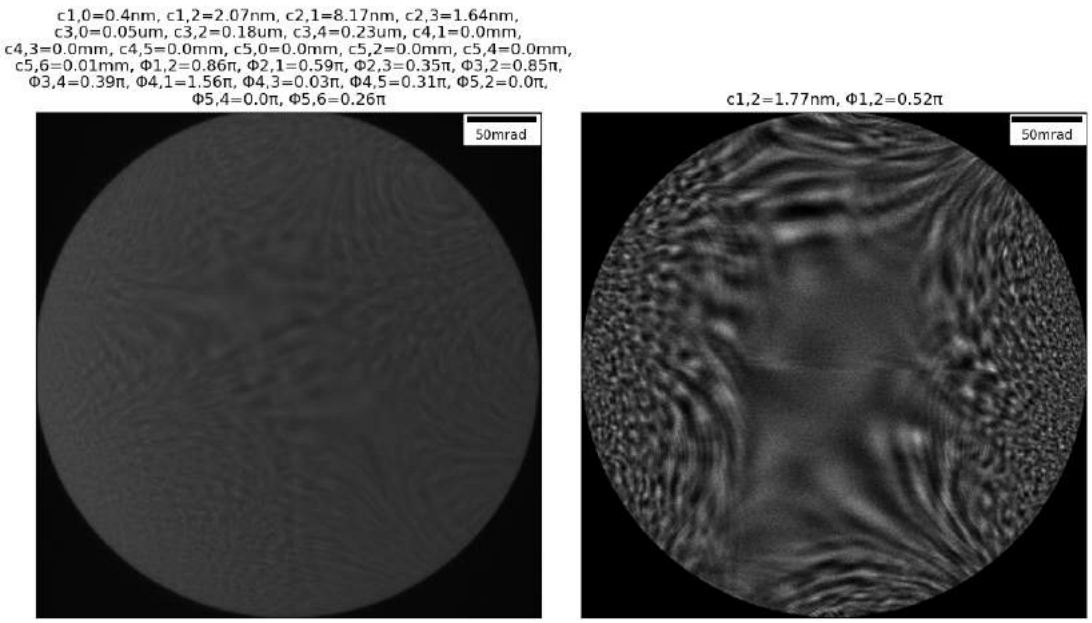


Figure 72: Experimental Ronchigram 9 (left) vs. predicted Ronchigram (right) generated by latest C3,0-corrected STEM network.

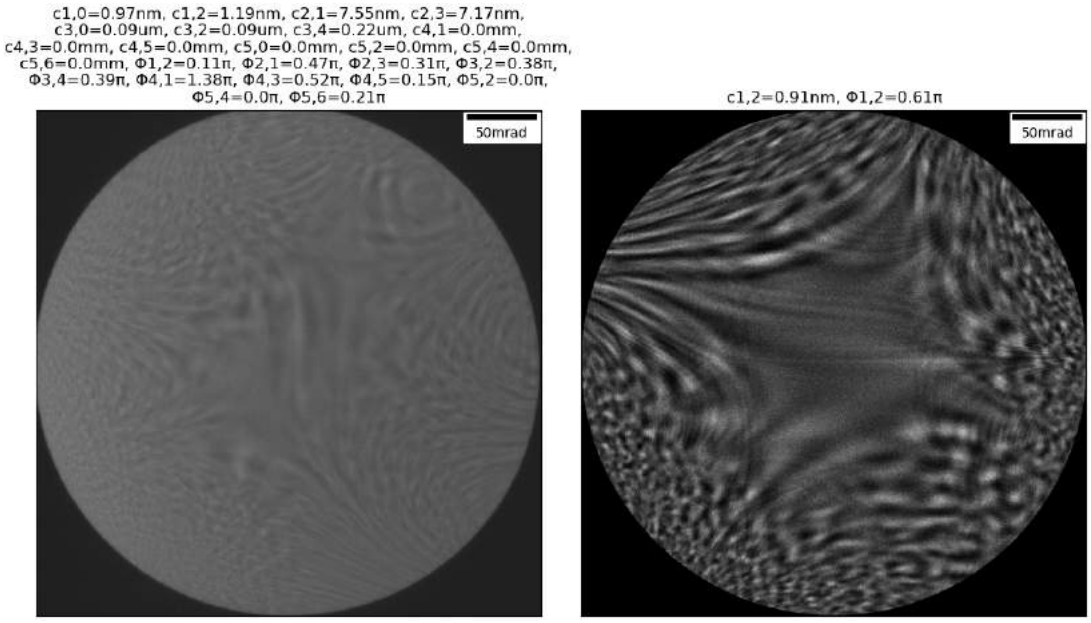


Figure 73: Experimental Ronchigram 10 (left) vs. predicted Ronchigram (right) generated by latest C3,0-corrected STEM network.

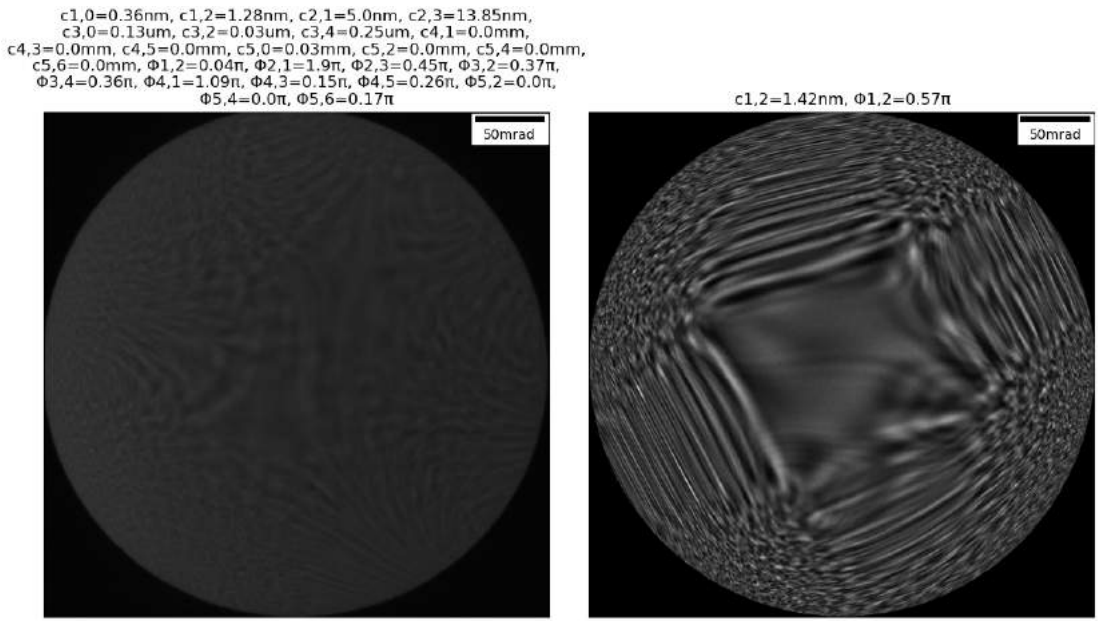


Figure 74: Experimental Ronchigram 11 (left) vs. predicted Ronchigram (right) generated by latest C3,0-corrected STEM network.

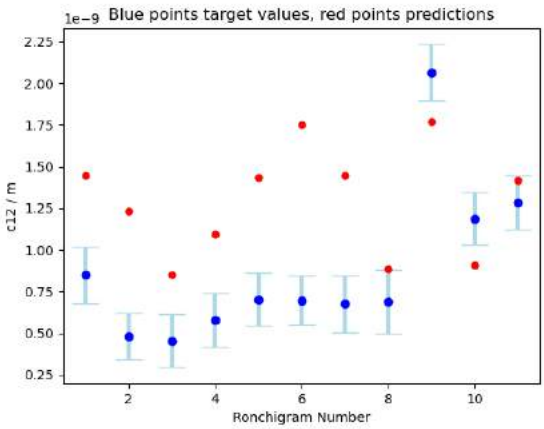


Figure 75: Plot of experimental Ronchigram c1,2 values estimated by JEOL COSMO™ with values predicted by latest C3,0-corrected STEM network.

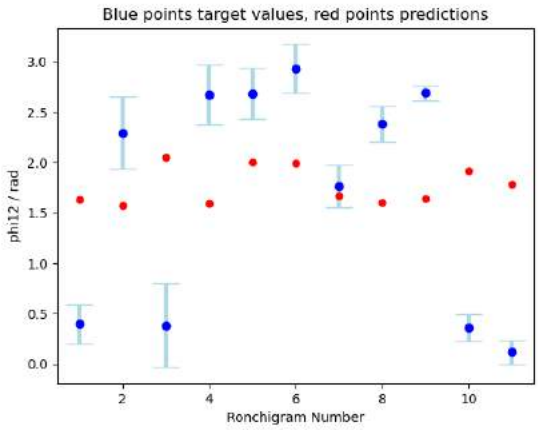


Figure 76: Plot of experimental Ronchigram ϕ1,2 values estimated by JEOL COSMO™ with values predicted by latest C3,0-corrected STEM network.

with the reconstructive inaccuracy using the JEOL COSMO™ aberration values leads to. Predicted simulations in in Figures 65, 66 and 72 (for experimental Ronchigrams 2, 3 and 9 respectively) show some semblance to experiment, but other pairs show little similarity besides having their central regions contained within similar dimensions. Furthermore, no discernible relationship between prediction success and illuminations (judged to decrease when quoted current and acquisition time decrease) seems present in these results, although this could be due to the inaccuracy of quoted currents, as mentioned in Section 6.5. In fact, prediction success is rare, with c1,2 predicted acceptably (firmly within the range of possibilities estimated in the experimental Ronchigrams by JEOL COSMO™) in only one case, specifically for Ronchigram 11. Additionally, for ϕ1,2 (see Figure 76), only 1 such result seems to be found, for Ronchigram 7. However, due to the very limited, and potentially biased, prediction spread here, this seemingly accurate prediction may only be by virtue of the fact that the Ronchigram 7's ϕ1,2 ground truth is within this spread—its error bar is approximately is nearly as wide as this range is. This alludes to an issue with the network in that it seems predisposed to predicting middling ϕ1,2 values, despite significant variations in ϕ1,2 quoted for experimental Ronchigrams. For future work, it may also be worth amending the network's tendency to predict mid-range ϕ1,2 values like in Figure 76–this may be caused by deficient network complexity, which can lead to some layers of the CNN outputting the same activation metrics for varying inputs, biasing the prediction made at deeper nodes and at the FC layer[56].

Although this is an ambitious test of network robustness, lessons may be taken from these results to improve performance on them. Specifically, these include the importance of matching specimen thickness to simulated projected potential–see the discrepancy within even the best comparison, that between Figures 50 and 51. Better matching of simulated projected potential to experiment could also involve considering how a thin, amorphous carbon film's transmission function would depend on frequency in reciprocal space, rather than simply using white noise without considering the varying nature of the specimen at different reciprocal space locations[57]. Experimental Ronchigrams were pre-processed before input to the CNN for inference via the same method by which all simulations were pre-processed for training and testing (see Section 6.2), so it is likely that the aforementioned details are responsible for great dissimilarity of what the network sees of the experimental Ronchigrams (Figure 77 for example) from what it is used to seeing, for

example, Figure 26. With what the network sees of experimental Ronchigrams appearing too unlike anything it has seen before, the network has not learnt to accurately estimate the values of the aberrations they feature.

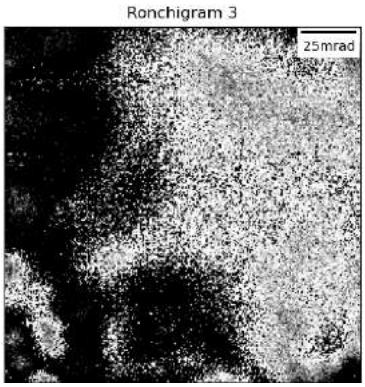


Figure 77: Experimental Ronchigram 3 pre-processed. Scale bar is from estimate of experimental Ronchigrams dimensions from comparing Figures 50 and 51. The counterpart in Figure 26 is for a simulation recreating experimental Ronchigram 3.

8 General Discussion and Conclusions

8.1 Conclusion

This project's ultimate aims were to use machine learning to achieve real-time aberration measurement by, at high speeds and accuracies, using the Ronchigram to evaluate magnitudes and angles of low-order aberrations present in the objective lens of the C3,0-corrected STEM, for varying illuminations and specimen types. More work is needed, particularly in training to improve accuracies at decreased experimental illuminations.

However, progress that hadn't been achieved pre-project was made in evaluation C1,2 in Ronchigram simulations emulating the C3,0-corrected STEM at 300kV, for high simulated electron currents of 100pA-1nA and acquisition times of 0.1s-1s. In validation, C1,2's magnitude and angle were evaluated by the EfficientNet-B2 with relatively low root mean square errors of 15.7% and 18.9% of the c1,2 and ϕ1,2 maxima in training, the network achieving a low mean inference time of 219 +/- 132 ms across 12 simulations in the Intel® Core™ i9-9980XE CPU @ 3.00GHz $\times$ 36. These accuracies and inference times may be suitable where quick, repetitive corrections are possible[25].

Over work completed before this project, this project has its successes:

1. Zemlin tableaus necessitate acquisition times of 30s and longer[12], while the CNN has been shown to achieve acceptable accuracies in simulations emulating Ronchigrams acquired in times of 1s and shorter, with an inference speed of 219ms, suggesting under 2s for one measurement of the C1,2 of an objective lens in practice.

2. No literature found at the beginning of this project had evaluated C1,2 from the Ronchigram using machine learning, but JEOL[25] trained a network to evaluate C2,1 to a precision (averaged across c1,2 and ϕ1,2) of 16% in experimental data, although specific details of this quantity are lacking. For the CNN in this thesis, root mean square errors were achieved for c1,2 and ϕ1,2 of 15.7% and 18.9% their values used for training, in the final validation stage for C1,2. However this value must be compared to that in tests with experimental Ronchigrams where aberration values are known to high accuracy and the specimen better matches simulation parameters than that mentioned in Section 6.5.

In work completed during this project by Bertoni et al.[12], C1,2 parameters in C3,0-

corrected STEM validation simulations was measured to a MAE of only about 4.8% of the maximum values of these parameters in simulations (100nm). In contrast, averaged over c1,2 and ϕ1,2 the final overall root MSE (computed instead because validation MAE was not saved) achieved by this project's network on C3,0-corrected STEM validation simulations was as high as 17.2% of the maxima of the C1,2 parameters shown to it during validation, although this may again be acceptable in the case of quick, repetitive correction. In [12], the network was trained for 200 epochs, whereas the network trained by the end of this project had been trained for a total of 89 epochs with c1,2 validation loss continuing to decrease but little time left. Moreover, in [12], the Ronchigrams were segmented in order to apply an astigmatism fitting routine in a method similar to that of the Zemlin tableau, as was done in [35]. It may be the prolonged training and more extensive data pre-processing that permitted such low errors to be achieved, methods that could be recreated in future work.

To conclude, although networks trained in this project may not be practicable themselves yet, the networks can be further trained to be acceptable methods of real-time aberration measurement; in this, network weights already achieved and findings like the best areas of the Ronchigram to show to the network in training could continue to be of assistance. Finally, the ability of the network to recognise two-fold astigmatism to an acceptably high accuracy now permits investigation into comparing methods like autocorrelation and machine learning in recognising C1,2 at decreasing levels of electron illumination.

8.2 Future Work

Work that might be worth undertaking in future includes:

1. Better fine-tuning of experimental Ronchigram acquisition parameters to simulations or vice-versa, to gain a fair test of the results of this thesis' methodology on real Ronchigrams from the C3,0-corrected STEM. As stated in Sections 6.1.2, this could involve simulating chromatic aberration, Gaussian circuit noise and impulse noise, and better tuning projected potential parameters to simulation as described at the end of Section 7.3.2.

2. Finding out at what levels of electron illumination the trained networks more accu-

rately evaluate C1,2 than commonplace methods like auto-correlation.

3. Undertaking further training to recognise C1,2 in different experimental setups, by using Ronchigrams simulating decreased levels of electron illumination and, for example, crystalline specimens. In training, regarding illumination, reconstructing low-illumination training simulations testing simulations and and experimental Ronchigrams could be done to facilitate network performance, using tools like a masked auto-encoder[58]. Moreover, it could be useful to better control the *diversity* of Ronchigrams shown to the network in training by selecting aberration combinations to be as dissimilar as possible, but sampling randomly from sub-ranges at fixed frequencies, thus ensuring not to bias the network by showing it aberration values all separated by fixed increments.

4. When the training data no longer bottlenecks network performance, it could be worth using methods like Bayesian optimisation to find machine learning parameters, including for hyper-parameters like learning rates, Nesterov momentum and weight decay, optimal for training. It could also be worth seeing if a larger version of EfficientNet (which might also solve some possible network complexity issues mentioned at the end of Section 7.3.2) leads to better increases in prediction performance per epoch. One might also compare the performances of machine learning methods outside the CNN, such as the k-nearest neighbours algorithm, the extreme learning machine or the mutlilayer perceptron, all of which have shown success in aberration correction in computer vision in the optical microscope[59], although outside of electron microscopy and without the Ronchigram. Finally, one could also trial reinforcement learning in tandem with aberration measurement from the Ronchigram, which has been successful in [33].

5. Finally, beyond training to recognise C1,2, it would be important to train the network to evaluate other aberrations–resolutions achievable are limited if only C1,2 is being controlled accurately.

9 Ethical and Sustainability Considerations

Ultimately, real-time aberration measurement from the Ronchigram by training CNNs to perform with varying illuminations and specimen types should not be a method that necessitates extra material–the tool will be software, which should be usable with a typical CPU, an object that research groups, including the one in which this Part II project was completed, tend to have at their disposal in great numeracy. However, on the way to suitably-trained neural networks for this process, there is still training that has been done in this project and that should be done in continuation of it, training whose implications still form a subsection of ethical and sustainability considerations that should be addressed.

Despite the relative eco-friendliness trained networks have in inference at this stage (where only one Ronchigram processing should occur every 2s at this stage in practice, given the 1s acquisition time that is typically done, as in Section 6.1.2, and the 219ms inference time as in Section 6.1.2), training has a relatively high computational demand per unit time, since approximately 40 Roncigrams were processed every 2 seconds in the final training run indicated in Figure 55. Since this training run was continued for 9s, and was only one of many done in this project, the computational demand in this project has mainly constituted training. Training in this project, and in future work for example in network optimisation, could be considered quite electrical demanding, ultimately contributing to CO2 production via electricity generation, thus contributing to global warming. In this project, computational demand was reduced via efforts like reducing unnecessary training, by: using early stopping, which detects using validation loss when a network is overfitting and no longer successfully being trained, thus halting the training process; and by the nature of PyTorch functions which cease to operate when training losses explode towards infinity (clearly signalling poor learning), causing gradients in gradient descent that are too large numerically to be handled, and thus training again is terminated. However, in future, training could be run in cloud computing services with servers based in regions colder than those of researcher training networks for aberration measurement from the Ronchigram; this could entail cloud providers like Amazon Web Services who have data centres in Scandinavia and Canada (Figure 78), regions typically colder than many more southerly regions from which network training

may be undertaken. However, when the networks trained are suitable for use in real-time aberration measurement from the Ronchigram, the time they save per scientific investigation in the STEM could compensate for the energy demands of training and inference by reducing the energy the STEM consumes (for generating, via super-cooling, temperatures as low as -30C that may be required for typical electron microscopy–see [60]) per scientific discovery, ultimately. According to another member of the Oxford Electron Image Analysis Group, 50% of microscopy time may need to be devoted by an experienced post-doctoral researcher on aberration measurement and correction without the benefit of real-time applicability, due to needs to replace specimens and adjust electron currents for current methods using the Ronchigram.

Figure 78: Map of AWS availability regions, blue markers indicating where regions currently exist. Figure comes from [61].

As explained, no new CPU or GPU was necessitated for this Part II project to be done, and nor should such apparatus need to be ordered by typical research groups furthering this work for themselves–if a GPU is too small for a typical network, a network with a smaller model size may be used, with the potential of achieving high enough accuracies to be acceptable in quick, repetitive corrections. Even if a new CPU is warranted for operation of the networks in tandem with STEM, the limited number of STEMs (see Section 4.4) means a low demand on new CPUs anyway, meaning this work should be sustainable regarding acquiring apparatus if necessary. Because of the negligible cost of new apparatus and the fact that the creation of networks involves developing software, new electronics generally aren't strictly required for this method to be used, meaning the specifics of waste electrical and electronic equipment recycling and its own implications (see [62]) should not be a concern.

Another implication of this work worth considering is the potential of this work to aid uses of STEM that may be considered unethical, for example in ballistics and in animal testing. However, the good that this method could do in enhancing pro-bono fields of research like those in medicine might offset these concerns. Moreover, animal testing typically involves acquiring biological material from animals and analysing it–eventually, the current aberration measurement method aims to be usable at low levels of electron illumination where biological sample damage isn't as likely, meaning less replacement biological material may be needed per investigation in this regard. However, it is a possibility that more investigations being permitted to occur in unit time by real-time aberration measurement means biological material is demanded more often; however, animal testing to the aim of cosmetic purposes, at least, is illegal in many countries[63], and the methodology developed in this project doesn't necessitate any biological material.

A more minor issue might be the umbrella concern of AI-enabled automation replacing human labour, thus depriving people of jobs. However, machine learning for aberration measurement and correction simply aims to replace current methods already being used by microscopists, not the microscopists themselves–the microscopists would of course be those initiating and directing the investigation, the neural networks would just be facilitating this. Regarding the facilitation of microscopy by neural networks trained to measure aberrations from the Ronchigram, given the time that existing practicable aberration measurement methods take, the aimed-at method may reduce the tedium involved in microscopy, thereby motivating microscopists, proliferating work and allowing breakthroughs to be made for the good in fields like cancer research–thus, continuing this work towards machine learning for real-time aberration measurement may indeed be an ethical step to take.

10 Project Management

10.1 Initial Ultimate Objectives and Milestones

As stated in PMF (Project Management Form) 1, the ultimate objectives were initially to "build" (set up and train) a CNN to accurately evaluate aberration constants in microscopes generating the Ronchigram and maximise the CNN's computational efficiency, by maximising inference speed for example. Due to lacking scientific knowledge, specific aberrations and microscopes weren't stated, nor was the meaning of "accurately". Although these aims weren't achieved as soon as planned in Gantt charts in the PMFs, besides maximising CNN computationally efficient (decided not to be a primary objective given EfficientNet-B2's sophisticatedly-designed performance, explained in Section 6.2), they were eventually: a CNN was trained to evaluate the magnitude and angle of an aberration, C1,2, in simulated Ronchigrams emulating the C3,0-corrected STEM, at an accuracy that may be acceptable where quick, repetitive corrections are available.

In achieving the aforementioned aims, it was planned (see PMF 1) to pass the following milestones:

1. Establishing a CNN for training to evaluate aberrations in Ronchigrams, established by coding pipelines to permit this;

2. Simulating Ronchigrams, done although these could have been made more experimentally realistic by simulating more experimental artefacts mentioned in Section 6.1.2;

3. Undertaking training, done using the simulations; and

4. Testing the CNN on experimental Ronchigrams, although the test was too ambitious to fairly represent the CNN's experimental capabilities, meaning further testing is warranted before continuing to modify simulations.

Regarding the aim in PMF 3 to make things work at significantly varying and increasing detector noise levels, this was one the statement "depending on progress, further objectives may be decided" in PMF 1 encapsulated, along with ensure applicability to varying specimen types, these aims crucial in eventually achieving real-time aberration measurement. These aims weren't detailed in PMF 1 because it wasn't anticipated there

would be time to achieve them but they were aims for the future; as expected, time was indeed lacking. However, the sequential nature of the incremental development approach mentioned in Section 6.4 meant that superfluous aims didn't need to be detailed as early as PMF 1 for coordination with other tasks.

10.2 How Early Results Matched Hypotheses

Compared to more science-based projects where overarching hypotheses may be established and tested, this project was not begun with many grand hypotheses since the aim was more engineering based, regarding giving a CNN functionality to achieve aberration recognition; the main hypothesis was that a CNN could indeed reach objectives detailed in PMF 1. Section 7 indeed demonstrates this to be the case. Moreover, at points in the project where this was not working very well, plans of action simply involved varying aspects of training such as the dimensions of the simulated Ronchigrams shown to the CNN in training and testing in order to improve learning and inference respectively–in this way, significant alternative routes were not really taken during this project, since it was always the plan to iterate the methodology in order to see what worked well, as explained in Section 6.4.

However, one significant initial hypothesis was that, given the fact that every point in a Ronchigram should be affected by the aberration values featured by the Ronchigram, training might not vary so much whether a lot of the Ronchigram was exposed to the network or not. On the contrary, although training was done for quite a while prior, the main project breakthrough came in late Hilary, where limiting training to the central region of the Ronchigram was found to greatly enable the recognition of 1st and 2nd order aberration magnitudes and angles in simulations featuring the aforementioned aberrations alone. See Section 7.1 where this is detailed.

10.3 Mitigating Setbacks in the Project

In the early stages, several potential setbacks were prevented. A major example of a possible setback was the potential of damage to the group computer in some way causing a loss of data which was crucial to the project. To prevent this from being an issue, from early on, Git was used to upload code written to GitHub, this being done daily. As a result, if there were group computer shutdowns or damages, data would still be accessible

in the cloud, protected by a company that invested more in maintenance and security of their systems than the Oxford Electron Image Analysis Group had the resources to. On the other hand, although code was being backed up, network weights (defining a network's trained capabilities) were not backed up due to the lack of resources available to make this easily doable. However, among backed up code was a configuration file that could be re-run to reproduce the same training results, so that was at least sufficient to mean all network weights achieved were reproducible.

Furthermore, for this project, at most one GPU was provided by the Oxford Electron Image Analysis Group at any time, since other members of the group were using the group computer as well, in which there were only two GPUs. Therefore, at most one training run could be run at any given time. However, it was anticipated there might be cases in which simultaneous training was needed, if enhanced productivity was required to meet project aims for example. If this happened to be of concern, it was planned to use funds allocated to Part II students to purchase the highest level of subscription to Google Colab, a service that could allow machine learning tasks like training of neural network to be undertaken in the cloud rather than locally. Such a subscription would have enabled use of the fastest, highest-memory GPUs available, at the longest times available per session (necessary since training runs could have taken several days).

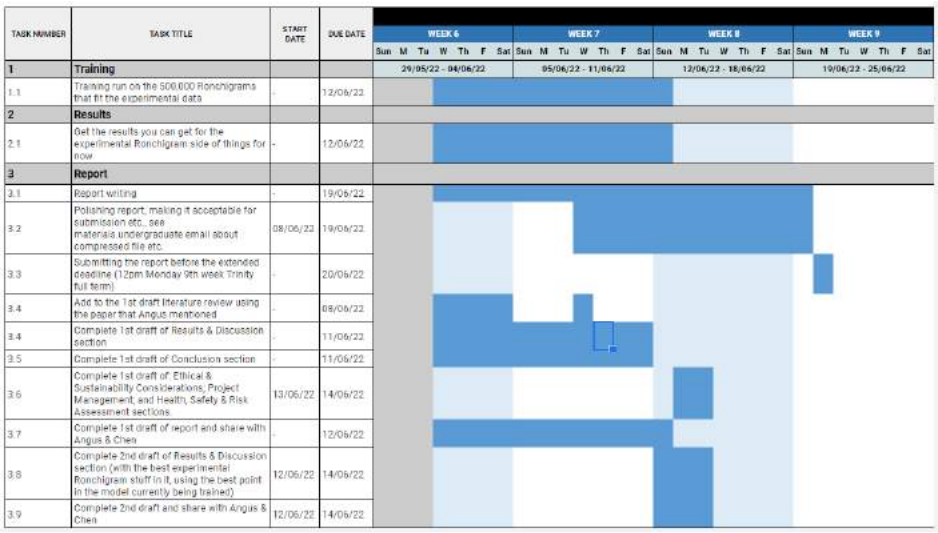
Regarding setbacks that did occur, the main ones were contracting COVID-19. This first happened early in Hilary Term, as mentioned in PMF 3, hindering my ability to work and leaving me with after-effects for weeks afterwards. No real plans could be made to avoid this, however, it didn't really cause any issues to the project besides ultimately reducing the final result of training regarding accuracy, for example. Fortunately, this wasn't a collaborative project so my absence had little effect on the work of other members of the research group. COVID-19 was contracted again in Trinity Term, affecting thesis writing. However, the The Proctors' Office granted me a 2-week extension which compensated for this setback.

10.4 Managing Resources

Regarding group computer GPU usage, a spreadsheet was established to efficiently share the GPUs between group members. Moreover, when it came to acquiring experimental Ronchigrams from the STEM at ePSIC as planned in PMF 1, it was decided this would

be done when the CNN had reached a suitable accuracy with simulations, especially since the nature of the Ronchigrams to be acquired would depend on the the experimental parameters for which the final CNN was trained, the status of the final CNN of course unknown early in the project. However, this happened to be in Trinity Term, so that was when acquisition was done by Dr. Huang, to whom booking the STEM was left.

Regarding the resource of time, extra Gantt charts were developed to manage time in the long-term (see Figure 79 for an example); to manage time in the short-term, daily to-do lists were made (see Figure 80 for an example) rather than daily timetables since it wasn't always predictable how long tasks would take, especially given that new coding functions like those in Matplotlib needed learning before use.



TASK NUMBER	TASK TITLE	START DATE	DUE DATE
1	**Training**		
1.1	Training run on the 500,000 Ronchigrams that fit the experimental data	-	12/06/22
2	**Results**		
2.1	Get the results you can get for the experimental Ronchigram side of things for now	-	12/06/22
3	**Report**		
3.1	Report writing	-	19/06/22
3.2	Polishing report, making it acceptable for submission etc., see materials,undergraduate email about compressed file etc.	08/06/22	19/06/22
3.3	Submitting the report before the extended deadline (12pm Monday 9th week Trinity full term)	-	20/06/22
3.4	Add to the 1st draft literature review using the paper that Angus mentioned	-	08/06/22
3.4	Complete 1st draft of Results & Discussion section	-	11/06/22
3.5	Complete 1st draft of Conclusion section	-	11/06/22
3.6	Complete 1st draft of: Ethical & Sustainability Considerations; Project Management; and Health, Safety & Risk Assessment sections	13/06/22	14/06/22
3.7	Complete 1st draft of report and share with Angus & Chen	-	12/06/22
3.8	Complete 2nd draft of Results & Discussion section (with the best experimental Ronchigram stuff in it, using the best point in the model currently being trained)	12/06/22	14/06/22
3.9	Complete 2nd draft and share with Angus & Chen	12/06/22	14/06/22

Figure 79: Part of the Gantt chart used in Trinity Term.

Regarding educational resources, well-documented software like Python were learnt online, while postgraduate lectures in electron microscopy were attended at the department.

10.5 Project Management Learning Experiences

Although most of the thesis was written in Trinity, it would have been helpful to devote the entire year to writing it as results came along, besides just preparing to write sections like the literature review by making notes on research papers throughout the year. However, this was difficult because of motivations to focus wholly on the project work itself, especially given that (despite efforts and planning) this was often a struggle. In hindsight, I would have leaned more on my supervisor and other members of the Oxford Electron

Things to do today:

- Today I must:
  - DONE Write the Ethics & Sustainability section
  - Write the Project Management section
  - Write the Risk Assessment section
  - Complete the experimental Ronchigram acquisition bit
  - Correct the report based on Chen's comments on it

1. DONE Decide which order to write the Ethics & Sustainability, Project Management and Risk Assessment sections
   a. Part II handbook has a default structure of Project Management, Ethics & Sustainability, and Risk Assessment
   b. Verdict: I am going to use the order I want to use

2. DONE Write the Ethics & Sustainability section in https://drive.google.com/drive/folders/1a7tqOMbETuaaqBiSEXj71KKz_x-lm7Q
   a. DONE Make notes about this section in https://docs.google.com/document/d/1PGA9qkTKvYMLVWnXtptmUUptLXraXBpWUEnzdy4Y/edit, including notes on:
      i. DONE The Part II handbook
      ii. DONE Canvas resources
      iii. DONE Presentation you made in Hilary
   b. DONE Plan out this section in https://docs.google.com/document/d/11b0rplpcFMhdc8vlp4NM_E2vfMDD-3GhmTODuzvdaTk/edit
   c. DONE Write this section in https://docs.google.com/document/d/11b0rplpcFMhdc8vlp4NM_E2vfMDD-3GhmTODuzvdaTk/edit
   d. Put what you've written in https://docs.google.com/document/d/11b0rplpcFMhdc8vlp4NM_E2vfMDD-3GhmTODuzvdaTk/edit into the thesis, moving:
      i. DONE The text
      ii. DONE Figures
      iii. DONE Load references
      iv. DONE Clear up in-text and in-caption references to:
         1. DONE Figures
         2. DONE Sections
         3. DONE Citations

3. NOT FINISHED Write the Project Management section
   a. Make notes about this section in https://docs.google.com/document/d/1AACw4lothVpXfy1qQGu9QbmogR1TcFmKVDAh8Ok/edit, including notes on:
      i. DONE The Part II handbook
      ii. DONE Canvas resources
      iii. DONE PMFs 1, 2 and 3
   b. Put what you have written in https://docs.google.com/document/d/1AACw4lothVpXfy1qQGu9QbmogR1TcFmKVDAh8Ok/edit into the thesis, specifically in the sections titled "Compulsory Final Chapter of the Part II Thesis: Project Management and the Part II Thesis (max. 1500 words, excluding the PM forms)" and "My ideas"

Figure 80: An example of a daily to-do list made in this project.

Image Analysis Group for verification of my code and machine learning strategies sooner, to make project work more fruitful and free up focus for aspects like thesis writing and getting further in the project.

11 Risk Assessment

All project work was carried out on a laptop. While processed experimental Ronchigrams were acquired as detailed in Section 6.5, this was done by Dr. Huang, meaning there were no experimental hazards that needed to be mitigated regarding myself. However, over the course of a year-long project, working solely on a laptop can pose risks, for example RSI (Repetitive Strain Injury) from typing and muscular and skeletal damage from poor posture while working at a laptop. In order to prevent harm to the body from these risks, ergonomic workstations were utilised for the vast majority of project work, by doing the majority of laptop work at the Mansfield College Main Library, where there were chairs whose positions were adjustable and tables that could be raised and lowered, to ensure that hands, wrists, forearms and elbows were aligned with the keyboard and that the lower back was supported, as was necessary to most cautiously prevent RSI [64] and muscular and skeletal damage. Work was also done at one of the Oxford Electron Image Analysis Group offices at the Department of Materials, where there were adjustable office chairs that could achieve the aforementioned levels of ergonomic safety.

Besides bodily injury, extensive laptop work poses risks like eye strain. In this project, to prevent this, prescription glasses featuring anti-glare lenses were purchased at the beginning of the academic year, the 20-20-20 rule of looking at something about 20 feet away for 20 seconds every 20 minutes of work was followed, and it was ensured the monitor was kept at eye level in order to reduce strain associated with changing gaze directions[65].

A final concern was the impact of isolated laptop work on mental health. To mitigate this, work was instead done in pseudo-social environments like the aforementioned library and the research group office, daily exercise was done in the morning in order to experience the mood and focus benefits of the endorphins secreted at such times[66], and short, daily walks were taken in nature to clear the mind and prevent exorbitant mental predominance of work.

12 Bibliography

References

[1] Noah Schnitzer, Suk Hyun Sung, and Robert Hovden. Introduction to the Ronchigram and its Calculation with Ronchigram.com. *Microscopy Today*, 27(3): 12–15, 2019. ISSN 1551-9295. doi: 10.1017/S1551929519000427.

[2] H. Sawada, T. Sannomiya, F. Hosokawa, T. Nakamichi, T. Kaneyama, T. Tomita, Y. Kondo, T. Tanaka, Y. Oshima, Y. Tanishiro, and K. Takayanagi. Measurement method of aberration from Ronchigram by autocorrelation function. *Ultramicroscopy*, 108(11):1467–1475, 2008. ISSN 0304-3991. doi: 10.1016/J.ULTRAMIC.2008.04.095.

[3] Mark Winey, Janet B. Meehl, Eileen T. O'Toole, and Thomas H. Giddings. Conventional transmission electron microscopy. *Molecular Biology of the Cell*, 25 (3):319–323, 2014. ISSN 1059-1524. doi: 10.1091/mbc.e12-12-0863.

[4] Nature Education. The relative scale of biological molecules and structures. `https://www.nature.com/scitable/content/the-relative-scale-of-biological-molecules-and-14704956/`, 2022. [Online; accessed 12-June-2022].

[5] Naresh Marturi. *Vision and visual servoing for nanomanipulation and nanocharacterization in scanning electron microscope.* PhD thesis, Université de Franche-Comté, 2013.

[6] E.M James and N.D Browning. Practical aspects of atomic resolution imaging and analysis in STEM. *Ultramicroscopy*, 78(1-4):125–139, 1999. ISSN 03043991. doi: 10.1016/S0304-3991(99)00018-2.

[7] Stephen J. Pennycook and Peter D. Nellist. *Scanning Transmission Electron Microscopy.* Springer New York, New York, NY, 2011. ISBN 978-1-4419-7199-9. doi: 10.1007/978-1-4419-7200-2.

[8] O. Scherzer. ber einige Fehler von Elektronenlinsen. *Zeitschrift fr Physik*, 101 (9-10):593–603, 1936. ISSN 1434-6001. doi: 10.1007/BF01349606.

[9] John Rodenburg. The objective lens of a TEM, the heart of the electron microscope. http://www.rodenburg.org/guide/t700.html, 2021. [Online; accessed 12-June-2022].

[10] JEOL. Keywords "astigmatism". https://www.jeol.co.jp/en/words/semterms/search_result.html?keyword=astigmatism. [Online; accessed 12-June-2022].

[11] J. M. Rodenburg and E. B. Macak. Optimising the resolution of TEM/STEM with the electron ronchigram. *Microscopy and Analysis*, 90(5-7), 2002.

[12] Giovanni Bertoni, Enzo Rotunno, Daan Marsmans, Peter Tiemeijer, Amir H. Tavabi, Rafal E. Dunin-Borkowski, and Vincenzo Grillo. Near-real-time diagnosis of electron optical phase aberrations in scanning transmission electron microscopy using an artificial neural network. 2022. URL https://arxiv.org/abs/2204.11126. [Online; accessed 20-June-2022].

[13] Noah Schnitzer, Suk Hyun Sung, and Robert Hovden. Optimal STEM Convergence Angle Selection Using a Convolutional Neural Network and the Strehl Ratio. *Microscopy and Microanalysis*, 26(5), 2020. ISSN 1431-9276. doi: 10.1017/S1431927620001841.

[14] David G. Voelz. *Computational Fourier Optics: A MATLAB Tutorial*. SPIE, 2011. ISBN 9780819482051. doi: 10.1117/3.858456.

[15] JEOL. Keywords "ronchigram" | Keywords | Glossary of TEM Terms. https://www.jeol.co.jp/en/words/emterms/search_result.html?keyword=ronchigram, 2022. [Online; accessed 15-June-2022].

[16] P.E. Batson. Aberration correction results in the IBM STEM instrument. *Ultramicroscopy*, 96(3-4):239–249, 2003. ISSN 03043991. doi: 10.1016/S0304-3991(03)00091-3.

[17] Phung and Rhee. A High-Accuracy Model Average Ensemble of Convolutional Neural Networks for Classification of Cloud Image Patches on Small Datasets. *Applied Sciences*, 9(21):4500, 2019. ISSN 2076-3417. doi: 10.3390/app9214500.

[18] Savya Khosla. CNN | Introduction to Pooling Layer - GeeksforGeeks. https://www.geeksforgeeks.org/cnn-introduction-to-pooling-layer/#:~:

text=Why%20to%20use%20Pooling%20Layers,generated%20by%20a%20convolution%20layer, 2021. [Online; accessed 12-June-2022].

[19] UCSB. UCSB Science Line. http://scienceline.ucsb.edu/getkey.php?key=144#:~:text=The%20average%20length%20of%20a,meters%20(5%20feet)%20long., 2021. [Online; accessed 12-June-2022].

[20] Andreas Engel and Christian Colliex. Application of scanning transmission electron microscopy to the study of biological structure. *Current Opinion in Biotechnology*, 4(4):403–411, 1993. ISSN 09581669. doi: 10.1016/0958-1669(93)90005-H.

[21] Diana B. Peckys and Niels de Jonge. Liquid Scanning Transmission Electron Microscopy: Imaging Protein Complexes in their Native Environment in Whole Eukaryotic Cells. *Microscopy and Microanalysis*, 20(2):346–365, 2014. ISSN 1431-9276. doi: 10.1017/S1431927614000099.

[22] Leslie A. Pray. Discovery of DNA Structure and Function: Watson and Crick. *Nature Education*, 1(1):100, 2008.

[23] Ding Liu, Bihan Wen, Xianming Liu, Zhangyang Wang, and Thomas S. Huang. When image denoising meets high-level vision tasks: A deep learning approach. *arXiv preprint arXiv:1706.04284*, 2018.

[24] David B. Williams and C. Barry Carter. *Transmission electron microscopy : a textbook for materials science*. Springer Science+Business Media, LLC, New York, 1 edition, 1996.

[25] Ryusuke Sagawa, Fuminori Uematsu, Keito Aibara, Tomohiro Nakamichi, and Shigeyuki Morishita. Aberration Measurement and Correction in Scanning Transmission Electron Microscopy using Machine Learning. *Microscopy and Microanalysis*, 27(S1):814–816, 2021. ISSN 1431-9276. doi: 10.1017/S1431927621003226.

[26] Stephen Selipsky. Stephen Selipsky's Answer to "How many electron microscopes are there in the world?". https:

//www.quora.com/How-many-electron-microscopes-are-there-in-the-world, 2016. [Online; accessed 12-June-2022].

[27] Peter Ercius, Markus Boese, Thomas Duden, and Ulrich Dahmen. Operation of TEAM I in a User Environment at NCEM. *Microscopy and Microanalysis*, 18(4): 676–683, 2012. ISSN 1431-9276. doi: 10.1017/S1431927612001225.

[28] M. Haider, H. Müller, S. Uhlemann, J. Zach, U. Loebau, and R. Hoeschen. Prerequisites for a Cc/Cs-corrected ultrahigh-resolution TEM. *Ultramicroscopy*, 108 (3):167–178, 2008. ISSN 03043991. doi: 10.1016/j.ultramic.2007.07.007.

[29] P.W. Hawkes. The correction of electron lens aberrations. *Ultramicroscopy*, 156: A1–A64, 2015. ISSN 03043991. doi: 10.1016/j.ultramic.2015.03.007.

[30] F. Zemlin, K. Weiss, P. Schiske, W. Kunath, and K.-H. Herrmann. Coma-free alignment of high resolution electron microscopes with the aid of optical diffractograms. *Ultramicroscopy*, 3:49–60, 1978. ISSN 03043991. doi: 10.1016/S0304-3991(78)80006-0.

[31] R.R Meyer, A.I Kirkland, and W.O Saxton. A new method for the determination of the wave aberration function for high resolution TEM: 1. Measurement of the symmetric aberrations. *Ultramicroscopy*, 92(2):89–109, 2002. ISSN 03043991. doi: 10.1016/S0304-3991(02)00071-2.

[32] R.R. Meyer, A.I. Kirkland, and W.O. Saxton. A new method for the determination of the wave aberration function for high-resolution TEM.: 2. Measurement of the antisymmetric aberrations. *Ultramicroscopy*, 99(2-3):115–123, 2004. ISSN 03043991. doi: 10.1016/j.ultramic.2003.11.001.

[33] H. Akima, Y. Hirayama, and T. Yoshida. Automated Alignment Cs Correction System for STEM (II): Installation of Reinforcement Learning. n.d. URL https://www.microscopy.org/MandM/2011/abstracts/LB-30.pdf.

[34] Great Learning Team. Introduction to VGG16 | What is VGG16? - Great Learning. https://www.mygreatlearning.com/blog/introduction-to-vgg16/, 2021. [Online; accessed 12-June-2022].

[35] A.R. Lupini, P. Wang, P.D. Nellist, A.I. Kirkland, and S.J. Pennycook. Aberration measurement using the Ronchigram contrast transfer function. *Ultramicroscopy*, 110(7):891–898, 2010. ISSN 03043991. doi: 10.1016/j.ultramic.2010.04.006.

[36] Suk Hyun Sung and Noah Schnitzer. noahschnitzer/ronchigram-matlab/simulation/shifted_ronchigram.m - GitHub. `https://github.com/noahschnitzer/ronchigram-matlab/blob/master/simulation/shifted_ronchigram.m`, 2020. [Online; accessed 12-June-2022].

[37] M. Schowalter, J. T. Titantah, D. Lamoen, and P. Kruse. Ab initio computation of the mean inner Coulomb potential of amorphous carbon structures. *Applied Physics Letters*, 86(11):112102, 2005. ISSN 0003-6951. doi: 10.1063/1.1885171.

[38] A. Harscher and H. Lichte. Inelastic mean free path and mean inner potential of carbon foil and vitrified ice measured with electron holography. *ICEM 14*, pages 553–554, 1998.

[39] Samuel W. Hasinoff. Photon , Poisson noise, 2012. URL `https://people.csail.mit.edu/hasinoff/pubs/hasinoff-photon-2012-preprint.pdf`. [Online; accessed 20-June-2022].

[40] USGS. Impulse Noise | U.S. Geological Survey -USGS.gov. `https://www.usgs.gov/landsat-missions/impulse-noise`. [Online; accessed 12-June-2022].

[41] Torch Contributors. SGD - PyTorch 1.11.0 documentation. `https://pytorch.org/docs/stable/generated/torch.optim.SGD.html`, 2022. [Online; accessed 18-June-2022].

[42] Henrique Mendonca. EfficientNet CIFAR-10 Ignite - Kaggle. `https://www.kaggle.com/code/hmendonca/efficientnet-cifar-10-ignite/log`, 2018. [Online; accessed 18-June-2022].

[43] Torch Contributors. torch.utils.data - PyTorch 1.11.0 documentation. `https://pytorch.org/docs/stable/data.html#torch.utils.data.DataLoader`, 2022. [Online; accessed 18-June-2022].

[44] Torch Contributors. torch.utils.tensorboard - PyTorch 1.11.0 documentation. https://pytorch.org/docs/stable/tensorboard.html, 2022. [Online; accessed 18-June-2022].

[45] Mingxing Tan and Quoc V. Le. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *ICML*, 2019. URL https://arxiv.org/abs/1905.11946. [Online; accessed 12-June-2022].

[46] Meta AI Research. Image Classification on ImageNet. https://paperswithcode.com/sota/image-classification-on-imagenet, 2022. [Online; accessed 12-June-2022].

[47] Cyril Meyer, Veronique Mallouh, Daniele Spehner, Etienne Baudrier, Patrick Schultz, and Benoit Naegel. Automatic Multi Class Organelle Segmentation For Cellular Fib-Sem Images. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pages 668–672. IEEE, 2021. ISBN 978-1-6654-1246-9. doi: 10.1109/ISBI48211.2021.9434075.

[48] Torch Contributors. RandomAffine - Torchvision 0.12 documentation - PyTorch. https://pytorch.org/vision/stable/generated/torchvision.transforms.RandomAffine.html, n.d.. [Online; accessed 18-June-2022].

[49] Torch Contributors. ToTensor - Torchvision main documentation - PyTorch. https://pytorch.org/vision/main/generated/torchvision.transforms.ToTensor.html, 2022. [Online; accessed 18-June-2022].

[50] Torch Contributors. CenterCrop - Torchvision 0.12 documentation - PyTorch. https://pytorch.org/vision/stable/generated/torchvision.transforms.CenterCrop.html, n.d.. [Online; accessed 18-June-2022].

[51] Torch Contributors. Resize - Torchvision main documentation - PyTorch. https://pytorch.org/vision/main/generated/torchvision.transforms.Resize.html, n.d.. [Online; accessed 18-June-2022].

[52] Torch Contributors. torch.nn.functional.normalize - PyTorch 1.11.0 documentation. https:

//pytorch.org/docs/stable/generated/torch.nn.functional.normalize.html, n.d.. [Online; accessed 18-June-2022].

[53] C. Barry Carter and David B. Williams. *Transmission Electron Microscopy: Diffraction, Imaging, and Spectrometry*. Springer International Publishing, Switzerland, 3 edition, 2016.

[54] Earl J. Kirkland. On the optimum probe in aberration corrected ADF-STEM. *Ultramicroscopy*, 111(11):1523–1530, 2011. ISSN 03043991. doi: 10.1016/j.ultramic.2011.09.002.

[55] Jeremy Jordan. Setting the learning rate of your neural network. https://www.jeremyjordan.me/nn-learning-rate/, 2018. [Online; accessed 12-June-2022].

[56] Varun Ranganathan. Varun Ranganathan's Answer to "Why does my own neural network give me the same output for different input sets?". https://www.quora.com/Why-does-my-own-neural-network-give-me-the-same-output-for-different-input-sets/answer/Varun-Ranganathan-4, 2018. [Online; accessed 14-June-2022].

[57] B. O'Malley, I. Snook, and D. McCulloch. Reverse Monte Carlo analysis of the structure of glassy carbon using electron-microscopy data. *Physical Review B*, 57 (22):14148–14157, 1998. ISSN 0163-1829. doi: 10.1103/PhysRevB.57.14148.

[58] Zhiyao Luo and Zhiyuan Ding. Idea for how to facilitate machine learning in aberration measurement from the Ronchigram at low electron illuminations., 4 2022.

[59] Yuncheng Jin, Yiye Zhang, Lejia Hu, Haiyang Huang, Qiaoqi Xu, Xinpei Zhu, Limeng Huang, Yao Zheng, Hui-Liang Shen, Wei Gong, and Ke Si. Machine learning guided rapid focusing with sensor-less aberration corrections. *Optics Express*, 26(23):30162, 2018. ISSN 1094-4087. doi: 10.1364/OE.26.030162.

[60] Electron Microscopy Sciences. Coolstage for SEM, LV, or VP - Electron Microscopy Sciences. https://www.emsdiasum.com/microscopy/products/equipment/cooling_stage.aspx#:~:text=At%20300Pa%2C%20the%20specimen%20temperature,specimen%20structure%20can%20be%20minimized., 2021. [Online; accessed 16-June-2022].

[61] Amazon Web Services inc. AWS Global Infrastructure. https://aws.amazon.com/about-aws/global-infrastructure/, 2022. [Online; accessed 16-June-2022].

[62] Health and Safety Executive. Waste Electrical and Electronic Equiprment recycling (WEEE), 2022. URL https://www.hse.gov.uk/waste/waste-electrical.htm. [Online; accessed 16-June-2022].

[63] Cherry Wilson. Animal testing: Which ones are used in UK experiments? - BBC. https://www.bbc.co.uk/news/newsbeat-47800019, 2019. [Online; accessed 16-June-2022].

[64] Alana Biggers. Repetitive Strain Injury (RSI): Causes, Prevention, and More. https://www.healthline.com/health/repetitive-strain-injury, 2017. [Online; accessed 17-June-2022].

[65] Mayo Clinic. Eyestrain - Diagnosis and treatment - Mayo Clinic. https://www.mayoclinic.org/diseases-conditions/eyestrain/diagnosis-treatment/drc-20372403, 2020. [Online; accessed 17-June-2022].

[66] Lawrence Robinson, Jeanne Segal, and Melinda Smith. The Mental Health Benefits of Exercise - HelpGuide.org. https://www.helpguide.org/articles/healthy-living/the-mental-health-benefits-of-exercise.htm, 2021. [Online; accessed 17-June-2022].

13 Appendix

All code referred to in this thesis may be found at

https://github.com/JamesGJ5/Part-II-Hilary-Onward.

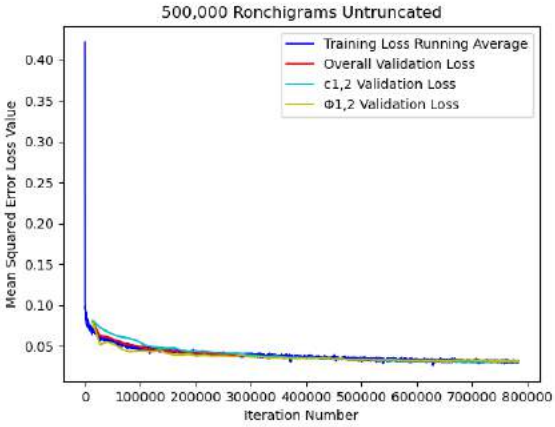


Figure 81: Figure 40 without truncation.