

Deep-learning in-situ classification of HIV-1 virion morphology

Juan S. Rey, Wen Li, Alexander J. Bryer, Hagan Beatson, Christian Lantz,  
Alan N. Engelman, Juan R. Perilla

PII: S2001-0370(21)00423-2  
DOI: <https://doi.org/10.1016/j.csbj.2021.10.001>  
Reference: CSBJ 1239

To appear in: *Computational and Structural Biotechnology Journal*

Received Date: 1 April 2021  
Accepted Date: 1 October 2021

Please cite this article as: J.S. Rey, W. Li, A.J. Bryer, H. Beatson, C. Lantz, A.N. Engelman, J.R. Perilla, Deep-learning in-situ classification of HIV-1 virion morphology, *Computational and Structural Biotechnology Journal* (2021), doi: <https://doi.org/10.1016/j.csbj.2021.10.001>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

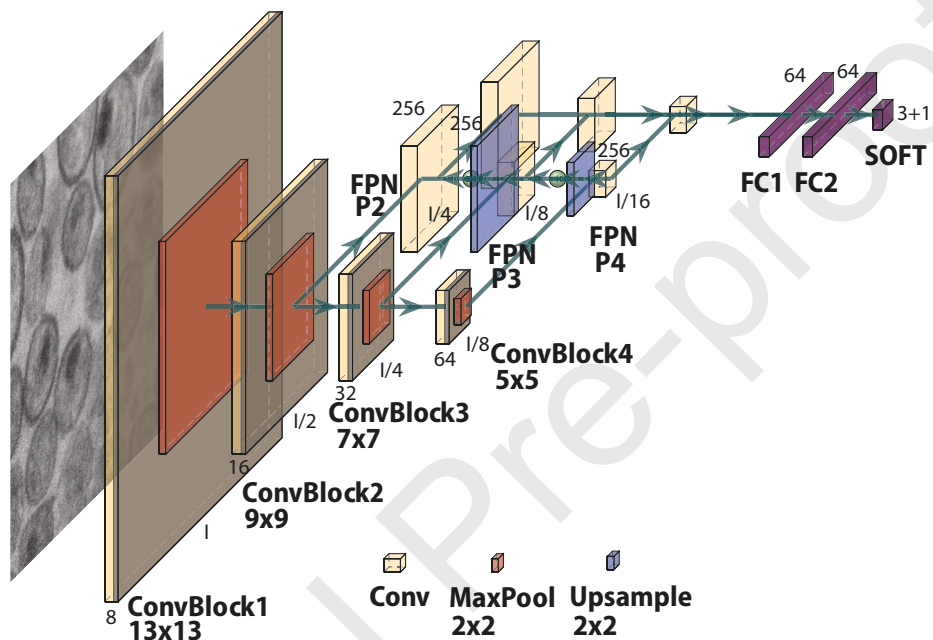
© 2021 Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology.



## Graphical Abstract

## Deep-learning in-situ classification of HIV-1 virion morphology

Juan S. Rey, Wen Li, Alexander J. Bryer, Hagan Beatson, Christian Lantz,  
 Alan N. Engelman, Juan R. Perilla



## Highlights

### **Deep-learning in-situ classification of HIV-1 virion morphology**

Juan S. Rey, Wen Li, Alexander J. Bryer, Hagan Beatson, Christian Lantz,  
Alan N. Engelman, Juan R. Perilla

- Efficient classifier capable of overcoming inherent problems of small-data training sets.
- Automated detection and classification of HIV-1 particle morphology from transmission electron micrographs.
- Three orders of magnitude speed increase in data processing with negligible loss in accuracy.

# Deep-learning in-situ classification of HIV-1 virion morphology

Juan S. Rey<sup>a,1</sup>, Wen Li<sup>c,d,1</sup>, Alexander J. Bryer<sup>a</sup>, Hagan Beatson<sup>a</sup>, Christian Lantz<sup>a</sup>, Alan N. Engelman<sup>c,d,\*</sup>, Juan R. Perilla<sup>a,b,\*</sup>

<sup>a</sup>*Department of Chemistry and Biochemistry, University of Delaware, Newark, Delaware, United States of America*

<sup>b</sup>*Data Science Institute, University of Delaware, Newark, Delaware, United States of America*

<sup>c</sup>*Department of Medicine, Harvard University, Boston, Massachusetts, United States of America*

<sup>d</sup>*Department of Cancer Immunology and Virology, Dana-Farber Cancer Institute, Boston, Massachusetts, United States of America*

---

## Abstract

Transmission electron microscopy (TEM) has a multitude of uses in biomedical imaging due to its ability to discern ultrastructure morphology at the nanometer scale. Through its ability to directly visualize virus particles, TEM has for several decades been an invaluable tool in the virologist's toolbox. As applied to HIV-1 research, TEM is critical to evaluate activities of inhibitors that block the maturation and morphogenesis steps of the virus lifecycle. However, both the preparation and analysis of TEM micrographs requires time consuming manual labor. Through the dedicated use of computer vision frameworks and machine learning techniques, we have developed a convolutional neural network backbone of a two-stage Region Based Convolutional Neural Network (RCNN) capable of identifying, segmenting and classifying HIV-1 virions at different stages of maturation and morphogenesis. Our results outperformed common RCNN backbones, achieving 80.0% mean Average Precision on a diverse set of micrographs comprising different

---

\*Corresponding authors.

*Email addresses:* `alan_engelman@dfci.harvard.edu` (Alan N. Engelman), `jperilla@udel.edu` (Juan R. Perilla)

<sup>1</sup>The authors wish it to be known, in their opinion, that the first two authors should be regarded as Joint First Authors.



experimental samples and magnifications. We expect that this tool will be of interest to a broad range of researchers.

*Keywords:* Quantitative biology, artificial intelligence, deep learning, electron microscopy, HIV-1, virology, computer vision

---

## 1. Introduction

### 1.1. *Electron microscopy and virus research*

Transmission electron microscopy (TEM) has long been used as a diagnostic tool in virology. Investigations of fluid samples from patients' skin lesions in the 1940s enabled the variola virus, which is the poxvirus that causes smallpox, to be discerned from the much larger varicella-zoster virus, which is a herpesvirus that causes chickenpox [1]. The introduction of negative stain materials, such as uranyl acetate and phosphotungstic acid, in the late 1950s, significantly improved ultrastructure resolution and thus was a springboard development for the use of TEM in modern day virology [2].

TEM has been invaluable to the discovery and diagnosis of many viral diseases that still plague the world today. For example, TEM data was instrumental in the initial classification of the AIDS virus, since named HIV-1 for human immunodeficiency virus 1, as a retrovirus [3]. TEM-based techniques are today used to diagnose pathologies associated with infection by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which is the cause of the worldwide COVID-19 pandemic (reviewed in [4]).

HIV-1 is taxonomically a lentivirus, which is one of six genera that comprise the Orthoretrovirinae subfamily of Retroviridae. As such, the viral structural proteins and replication enzymes are expressed in infected cells as Gag and Gag-Pol polyproteins that become cleaved by the viral protease (PR) enzyme during the process of virus maturation (reviewed in [5]). Typical mature HIV-1 particles, which are approximately 90 to 120 nm in diameter, harbor an internal core that is composed of a conical shell of capsid protein that houses the viral ribonucleoprotein complex (RNP) composed of two copies of the viral RNA genome, the structural protein nucleocapsid (NC), and reverse transcriptase and integrase (IN) enzymes [6] (Figure 1). The RNP is the most electron-dense component of HIV-1 particles [7]. In immature particles, which are non-infectious, the electron density presents as a toroidal structure in proximity to the viral membrane [7] (Figure 1).

TEM has been invaluable to studies of HIV-1 inhibitors that interfere with proper virion maturation. Compounds that inhibit HIV-1 PR activity block polyprotein processing and hence arrest HIV-1 replication at the maturation step (reviewed in [8]). A second class of HIV-1 maturation inhibitor, which is typified by bevirimat, binds to the protein substrate to inhibit the final cleavage of Gag processing between capsid and spacer peptide 1 (reviewed in [9]). Removal of the IN domain from the C-terminus of Gag-Pol can also increase the frequency of immature particles in HIV-1 virion preparations [10, 11]. IN missense mutations can moreover elicit eccentric HIV-1 particle formation, where the electron-dense RNP appears outside the viral core, often in association with the viral membrane [10, 11, 12, 13] (Figure 1). The allosteric IN inhibitor (ALLINI) class of preclinical HIV-1 compounds elicits eccentric particle formation (reviewed in [9]). In this way, the inhibitors hyper-multimerize IN to preclude its binding to RNA in the virus particle [14, 13].

### *1.2. Machine Learning applied to transmission electron microscopy*

Over the last five years, the application of Machine Learning (ML) in biomedical image processing has increased significantly [15]. For instance, image classification has enabled diagnostic prediction of Alzheimer’s disease in patients from brain MRIs [16] and SARS-CoV-2 detection from chest X-Ray scans [17]. For microscopy image analysis, where individual detection and classification of substructures of images are necessary, two main frameworks have been applied: object segmentation and object detection. While the first aims to classify the pixels in an image predicting the probability that they belong to a certain class, object detection uses a per-region approach for classifying object instances. These techniques have been applied for the detection of cancer cell nuclei [18], the segmentation of neural membranes [19], segmentation of feline calicivirus [20] and virus classifications [21]. So-called Convolutional Neural Networks (CNNs) have proven useful for the semantic segmentation of small extracellular vesicles (sEVs) from TEM micrographs [22, 23]. These approaches include U-Net [24], a CNN based on the combination of downsampling and upsampling layers with connections between the convolutional layers.

Although so-called deep learning models like Sparse Autoencoders (SAE) and Recurrent Neural Networks (RNNs) have been applied to medical imaging [25, 26], the most popular approach continues to be CNNs. The latter class of models have proven their usefulness in classification, detection and

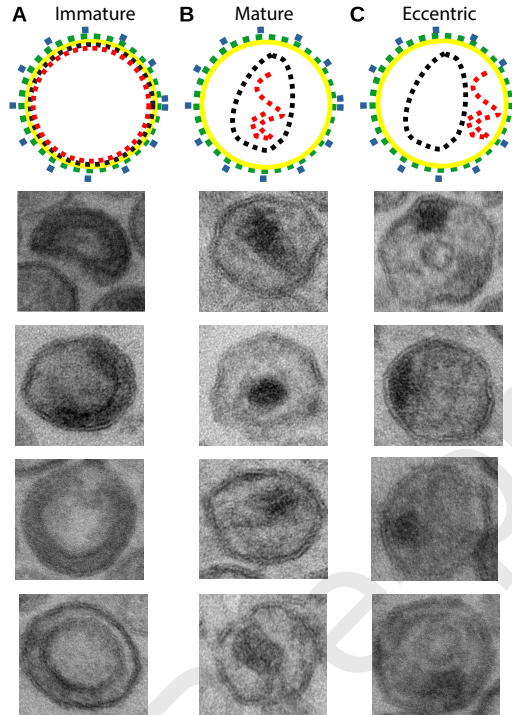


Figure 1: HIV-1 virion morphologies pertinent to this study. **A** Schematic representation showing the configuration of an immature virion and samples from TEM micrographs. The color scheme cartoons the following components from exterior to interior: blue, envelope glycoproteins; green, lipid bilayer; yellow, matrix protein; black, capsid protein; red, RNA. **B** Schematic representation showing the configuration of a mature virion and samples from TEM micrographs. **C** Schematic representation showing the configuration of an eccentric virion and samples from TEM micrographs.

segmentation tasks across a broad range of fields and applications, producing results up to par with medical experts [27, 28].

In the present manuscript, we present an end-to-end Deep Learning based method for the automated detection and classification of HIV-1 virion morphologies from input TEM micrographs. Our pipeline is composed of two main components, an object detection pipeline based on a Faster RCNN [29] architecture, and a stratification layer that can employ different classification backbones. The classification backbones supported include ResNets and a novel CNN, named *TEMNet*; the latter, designed and trained from scratch for the detection of HIV-1 virions from TEM micrographs. ResNets are pre-

trained using the ImageNet database, in our pipeline we support ResNet101, ResNet101v2 and InceptionResNetv2. Although ResNets are pretrained with photon-based photographs, here we show that both ResNet101, ResNet101v2 and Inception-ResNet-v2 accurately derive morphology distributions across mutant and clinical isolate samples after fine-tuning. In addition, our novel-classifier (TEMNet) proves to be a competitive classification backbone, trading off some accuracy (2% less accurate) for an 18% increase in speed and a 94% decrease in memory usage. Altogether, we find that our method is efficient and robust for *in situ* HIV-1 virion detection across different morphotypes, predicting statistical distributions agreeing with results from end-user visual inspection, while being up to three orders of magnitude faster. The application of our method to TEM micrographs achieved a mean Average Precision (mAP) of up to 80.0%.

Our implementation is based on the Tensorflow[30] Keras framework. The source-code has been made available through <https://github.com/Perilla-lab/TEMNet>.

## 2. Methods

### 2.1. Virus samples and TEM

All viruses analyzed in this study were generated from proviral DNA molecular clones. HIV-1 strain NL4-3 (HIV-1<sub>NL4-3</sub>) was generated from pNL4-3 [31] or pNL43/XmaI [32] while HIV-1 YU-2 and HIV-1 JR-CSF were generated from respective plasmids pYU-2 [33] and pYK-JRCSF [34]. Mutations in *pol* corresponding to IN changes D116N, N184L, and delIN, as well as PR active site mutation D25A, were introduced into pNL43/XmaI by site-directed mutagenesis using the primers listed below. The presence of desired mutations and absence of unwanted secondary changes were verified by Sanger sequencing. Previously described WT HIV-1<sub>NL4-3</sub> produced in the presence of dimethyl sulfoxide or the ALLINI BI-D, as well as IN mutant L241A, E96A, and N18I viral micrographs [7, 13, 35] were additionally used for RCNN training in this study.

Viruses were generated from plasmid DNAs by transfecting HEK293T cells, which were grown in Dulbecco's modified Eagle's medium supplemented to contain 10% fetal bovine serum, 100 IU/ml penicillin, and 100  $\mu$ g/ml streptomycin at 37°C in the presence of 5% CO<sub>2</sub>. Briefly, cells grown in two 15-cm dishes (10<sup>7</sup> cells per dish) were transfected with 30  $\mu$ g plasmid DNA using PolyJet DNA transfection reagent as recommended by the manufacturer

|         |        |  |
|---------|--------|--|
| D116N   | AE6540 | 5'-CCTTCTAAATGTGTACAA<br>TTAGCTGCCATATTCC-3'       |
|         | AE6541 | 5'-GTAAAAACAGTACATACA<br>AACAATGGCAGCAATTTTC-3'    |
| N184L   | AE7199 | 5'-GGCAGTATTCATCCACCTTT<br>TTAAAAGAAAAGGGGGGATT-3' |
|         | AE7200 | 5'-CCTTTTCTTTTAAAAAGG<br>TGGATGAATACTGCC-3'        |
| delIN   | AE4903 | 5'-CAGGAAAGTACTATTTTG<br>AGATGGAATAGATAAGGC-3'     |
|         | AE4904 | 5'-GCCTTATCTATTCCATCT<br>CAAAATAGTACTTTCCTG-3'     |
| PR D25A | AE7644 | 5'-GGAAGCTCTATTAGCTACAG<br>GAGCAGATGATAC-3'        |
|         | AE7645 | 5'-CTGTATCATCTGCTCCTG<br>TAGCTAATAGAGCTTC-3'       |

Table 1: Oligonucleotides used to introduce indicated changes into pNL43/XmaI DNA.

(SignaGen Laboratories). Two days after transfection, cell supernatants were filtered through 0.22  $\mu\text{m}$  filters and pelleted by ultracentrifugation using a Beckman SW32-Ti rotor at 26,000 rpm for 2 h at 4°C. Virus pellets were fixed with 1 mL fixative (2.5% glutaraldehyde, 1.25% paraformaldehyde, 0.03% picric acid, 0.1 M sodium cacodylate, pH 7.4) overnight at 4°C. The following steps were conducted at the Harvard Medical School Electron Microscopy core facility. Samples were washed with 0.1 M sodium cacodylate, pH 7.4, and postfixed with 1% osmium tetroxide and 1.5% potassium ferrocyanide for 1 h, washed twice with water, once with maleate buffer (MB), and incubated in 1% uranyl acetate in MB for 1 h. Samples washed twice with water were dehydrated in ethanol by subsequent 10 min incubations with 50%, 70%, 90%, and then twice with 100%. The samples were then placed in propyleneoxide for 1 h and infiltrated overnight in a 1:1 mixture of propyleneoxide and TAAB Epon (Marivac Canada Inc.). The following day, the samples were embedded in TAAB Epon and polymerized at 60°C for 48 h. Ultrathin sections (about 60 nm) were cut on a Reichert Ultracut-S microtome, transferred to copper grids stained with lead citrate, and examined in a JEOL 1200EX transmission electron microscope with images recorded on an AMT 2k CCD camera. Images were captured at 30,000 $\times$ , 25,000 $\times$  or

20,000 $\times$  magnification. Micrographs were stored on a 8-bit single-channel TIFF lossless format. In contrast to photon-based microscopy, where each pixel in the TIFF files encodes the wavelength of the photon, the TIFF files used in the present study contained electron intensities.

## 2.2. Data preparation

In order to build a robust neural network, capable of identifying HIV-1 virions across different experimental conditions we built training and validation datasets from micrograph samples using IN and PR mutant viruses to mimic eccentric and immature particle morphologies, respectively. In total, 59 micrographs imaged at 30,000 $\times$  magnification were assigned morphology labels. This dataset can be found freely at [10.5281/zenodo.5149062](https://zenodo.org/record/5149062).

The raw TEM micrographs were then pre-processed. First, the TEM micrographs were cropped, removing the image labeling information added by the microscope and standardizing the micrograph size to  $4,000 \times 2,620$  pixels. Additionally, since object detection tasks perform segmentation and classification of image-based objects by passing regions of interest through a convolutional network, bounding box coordinates  $(x_i, y_i, w_i, h_i)$  were assigned to each of the labeled virions in each micrograph.

A reasonably sized dataset is vital for training a deep neural network, which especially applies with CNNs, where the number of learnable parameters can reach millions and can quickly overfit if the number of training samples is too small. A known paradigm to solve this issue is via *transfer learning*, where a network is first pre-trained on a massive dataset like ImageNet [36] and then trained on the smaller target dataset. However, it has been shown that for object detection tasks [37], results on par with ImageNet pre-trained networks can be achieved when training from a random initialization (from scratch) with a dataset as low as 10k samples [38], given sufficient training time.

Two approaches were implemented to effectively increase the size of our data: First, each micrograph was cropped into overlapping regions of  $1,024 \times 1,024$  pixels. To generate the virion classes and box coordinates inside each cropped region, HIV-1 particles were counted as ground truth only if at least 75% of the area spanned by their respective bounding box was inside the cropped region. The features of the detected particles accordingly remained consistent across the datasets and misclassification noise, where a small section of a virion is mistakenly classified, was downplayed. This method generated between 1 to 48 regions where particles were present and



increased the number of images in the dataset to 2,730. The second approach consisted of applying offline augmentations to the cropped images as follows: each input image was transformed applying horizontal flipping, vertical flipping, 180° rotations and gaussian noise with a mean of 0 and a standard deviation of 1. This increased the dataset by a factor of four, generating new images and labels that were consistent with the features and morphologies of the HIV-1 virions in the micrographs, while the modifications on the images reduced overfitting on the training process. Together, the strategy described above, which is represented in Figure A.9 increased our dataset to 13,650 images. The latter dataset was divided for network optimization, yielding 10,725 images for training and 2,925 images for validation.

### 2.3. Region-based Convolutional Neural Networks (RCNN)

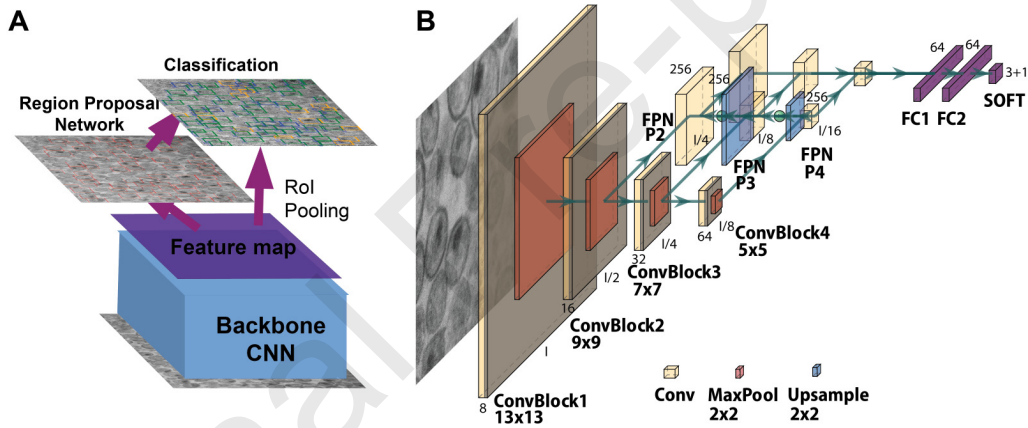


Figure 2: Network architecture of the deep-learning classifier developed in the present work. The methodology is built upon a two component classifier system that is able to segment the particles (RPN) and classify the segmented regions (CNN). **A** Schematic of a two-stage Faster RCNN architecture for multiple object detection and classification (instance segmentation). Faster RCNN uses a Region Proposal Network (RPN) to generate Regions of Interest (RoI) for classification. Both RPN and classifier heads share the same backbone CNN. **B** Our backbone CNN, TEMNet, is composed of several ConvBlocks (Convolution, GroupNormalization and ReLU activation) and MaxPooling layers. We used a Feature Pyramid Network (FPN) to generate multi-scale feature maps on which to generate predictions. Network activations were funneled to three output channels, via two fully-connected layers and soft max output. Each output channel denotes one classification of viral particle.

Developing an algorithm to identify and classify HIV-1 virions from TEM micrographs is in essence an object detection problem, where the goal is to

classify individual object instances in an image and localize each one using a bounding box. For this task we employed the *Region-based Convolutional Neural Network* [39] (RCNN) architecture.

RCNNs are based on applying a *Convolutional Neural Network* (CNN) to evaluate classification on a number of candidate Regions of Interest (RoI) delimited by bounding boxes. In this sense, RCNNs are two-stage object detection architectures since a network proposing the candidate RoIs is first necessary before the backbone CNN can be applied for classification. To efficiently generate RoIs, Ren, et.al [29] proposed the *Faster RCNN* architecture where, as shown in Figure 2(A), a *Region Proposal Network* shares the convolutional backbone used for classification and outputs a set of rectangular RoIs along with an *objectness score* indicating the probability of an object inside the RoI belonging to a class vs the background. The RPN works by generating *anchors*, i.e., sliding windows, of different sizes and scale ratios over the last Convolutional Feature Map output of the backbone CNN. Each anchor is mapped to an intermediate low dimension feature map and then connected to two fully connected layers, for regression of bounding box coordinates and computation of a class score that determines whether or not there is an object in the region. The RoIs proposed by the RPN are then passed through a RoI pooling layer [40] or RoIAlign layer [41] where their features are extracted via average or max pooling and then passed to the classifier heads. In this way, the final fully connected and softmax layers assign a per-class probability to each of the proposed RoIs.

Since the RCNN is a two-stage method that predicts both the bounding box localization  $(\hat{x}_i, \hat{y}_i, \hat{w}_i, \hat{h}_i)$  as well as the classification probabilities  $\hat{p}_i$  of object instances in an image with ground truth classes  $u_i$  and bounding box localization  $(x_i, y_i, w_i, h_i)$ , the error function to minimize network training is a multi-task loss consisting of two parts [40]:

$$L(\{\hat{p}_i, u_i\}, \{\hat{t}_i, t_i\}) = \frac{1}{N_{\text{class}}} \sum_i L_{\text{class}}(\hat{p}_i, u_i) + \lambda \frac{1}{N_{\text{RoI}}} \sum_i [u \geq 1] L_{\text{loc}}(\hat{t}_i, t_i) \quad (1)$$

where the normalizing parameters are the number of classes  $N_{\text{class}}$  and the number of regions of interest proposed  $N_{\text{RoI}}$ . The first loss

$$L_{\text{class}}(\hat{p}_i, u_i) = -\log \left( \frac{e^{\hat{p}_{i,u_i}}}{\sum_v e^{\hat{p}_{i,v}}} \right) \quad (2)$$

is the cross entropy log loss that the RoI proposal  $i$  belongs to the class  $u$  with a probability  $\hat{p}_{i,u}$ . While the second loss is calculated only when the



predicted region is not classified as background ( $u = 0$ ), as indicated by the Iverson bracket function  $[u \geq 1]$ , and it's given by a smooth L1 loss

$$L_{\text{loc}}(\hat{t}_i, t_i) = \sum_{c \in \{x, y, w, h\}} \text{smooth}_{L1}(\hat{t}_{i,c} - t_{i,c}) \quad (3)$$

with the coordinate offsets  $t_c$  defined as

$$\begin{aligned} \hat{t}_x &= (\hat{x} - x_a)/w_a, & \hat{t}_y &= (\hat{y} - y_a)/h_a, \\ \hat{t}_w &= \log(\hat{w}/w_a), & \hat{t}_h &= \log(\hat{h}/h_a), \\ t_x &= (x - x_a)/w_a, & t_y &= (y - y_a)/h_a, \\ t_w &= \log(w/w_a), & t_h &= \log(h/h_a), \end{aligned} \quad (4)$$

for the coordinates  $\hat{x}$ ,  $x$  and  $x_a$  denoting the predicted bounding box, ground truth box and anchor box coordinates, respectively. The hyperparameter  $\lambda$  is a weight that controls the balance between the two tasks error. As in [40], we use  $\lambda = 1$  for the training procedure.

For our purposes, the Faster RCNN architecture is incorporated for object detection tasks which allows experimentation by letting the user implement different backbone CNNs to be used for both RoI proposal and final object classification as shown in Figure 2 (A). In the present paper, we implemented the Residual Convolutional Neural Network *ResNet101* [42] architecture as well as its variation *ResNet101v2* [43] and the residual inception network *Inception-ResNet-v2* [44] along with our own compact convolutional backbone architecture named *TEMNet*.

#### 2.4. TEMNet

As shown in Figure 2(B) TEMNet is a sequential architecture composed of four convolutional blocks and max pooling layers. Each convolutional block consists of a 2D convolution followed by a normalization and a ReLu activation. Because convolutional blocks use padding to conserve the tensor size of the previous feature map, the feature map size is reduced by a factor of only 1/2, applying a max pooling layer of kernel size 2 after each convolutional block. Convolutional blocks 1, 2, 3 and 4 use kernel sizes of  $13 \times 13$ ,  $9 \times 9$ ,  $7 \times 7$  and  $5 \times 5$  respectively; sequentially decreasing in order to adapt to the reduced feature map size after max pooling on each one of the networks stages. As a way to mitigate overfitting, we added a Gaussian noise layer with standard deviation of 0.1 after the first max pooling layer, to act as a regularization layer for training while being inert for inference.

Normalization is essential for convergence of a deep network during training. However, batch normalization requires a sufficiently large batch size [45] which is not available in object detection tasks where a small batch size is necessary to keep a high image resolution. In this case, batch normalization can lead to inaccurate batch statistics.

On the one hand, pretraining and transfer learning are crucial techniques to facilitate the convergence of the network loss. In the case of ResNet backbones, weights trained on ImageNet [36], a massive-scale image classification database are readily available online. These pretrained weights were used as a starting point for training on our dataset where batch normalization layers were frozen, effectively transforming them to linear layers and the batch statistics learned on the massive-scale dataset were transferred to the new network. Results with and without pretraining for ResNet101 are shown in Figure A.10.

On the other hand, normalization on a network trained from scratch cannot benefit from transfer learning. Instead, for TEMNet we implemented group normalization [46], which normalizes along the channel axis instead of the batch axis.

In order to generate multi-scale feature maps on which to generate predictions, we used a Feature Pyramid Network [47] (FPN) for both our TEMNet and ResNet backbones. The ResNet implementation was done according to the original FPN paper [47] and zero padding was added to make the layers of Inception-ResNet-v2 compatible; while for TEMNet, in a similar manner the output of each of the max pooling layers  $\{C1, C2, C3, C4\}$  was used to generate the pyramid feature maps  $\{P2, P3, P4\}$ . For this procedure, every layer was passed through a  $1 \times 1$  convolution to standardize the number of filters (256), this convolution is known as a lateral connection. The top-down pathway was then built, starting with the coarsest resulting feature map  $P4$  (generated from  $C4$ ). The latter was upsampled with a  $2 \times 2$  kernel and added to the underlying  $C3$  feature map to generate  $P3$ , afterwards  $P3$  itself was upsampled and added to  $C2$  to generate the feature map  $P2$ . Finally, a  $3 \times 3$  convolution was applied on each feature map  $\{P2, P3, P4\}$ . These feature maps work as pyramid “levels” to which RoIs were mapped according to their size. Specifically, following [47] a RoI with height  $h$  and width  $w$  will be assigned to the pyramid level  $P_k$

$$k = \left\lfloor 4 + \log_2(\sqrt{wh}/2620) \right\rfloor \quad (5)$$

where 2,620 is the number of pixels constituting the smaller side of a micrograph and 4 is the single scale level for a RoI. Predictions on each of the pyramid levels were then funneled to two fully connected layers with 64 neurons for TEMNet and 1,024 for ResNet and Inception-ResNet, then finally to a softmax layer where per-class probability was assigned on 3+1 channels: three for our virion classifications: eccentric, mature, immature; and one for background.

For the training procedure, ResNets were initialized from ImageNet pre-trained weights and then fine-tuned on a dataset of 1,806 isolated HIV-1 virion samples. TEMNet was trained from scratch on the same dataset. Afterwards, the CNN backbones were initialized on their fine-tuned weights and trained individually on the RPN network for RoI proposal generation for 50 epochs on the cropped micrograph dataset consisting of 13,650 images. Finally, the RPN trained weights were used as initialization for training the CNN backbones on the full RCNN architecture for 50 epochs using the cropped micrograph dataset.

The input image was resized to  $512 \times 512$  pixels allowing a batch size of 8 images on a single NVIDIA V100 GPU. Stochastic Gradient Descent (SGD) was used to train the model with a starting learning rate of 0.01 which decreased by a factor of 10 every time a learning plateau was encountered on the validation loss. Weight decay was set to 0.0001, learning momentum to 0.9 to avoid the training getting stuck on a local minimum and gradient clipping norm to 5.0 to avoid exploding gradients. We used 50 RoIs per image for training and 20 for validation. Training Faster RCNN with TEMNet took less than 9 hours on our cropped dataset. Training error with ResNet and TEMNet backbones are shown in Figure 3. From this training the weight checkpoint, which achieves the lowest validation error, was chosen to avoid overfitting (Early stopping). The feature map activations learned by the convolutional blocks and FPN levels of TEMNet are presented in Figure 4

### 3. Results

#### 3.1. Prediction generation approach

As a result of training the network on a dataset composed of micrograph croppings, our Faster RCNN network can generate predictions on  $1,024 \times 1,024$  croppings of TEM micrographs. In order to generate end-to-end predictions on raw TEM images we devised a method to segment a micrograph via a sliding window. As illustrated in Figure 5, we scanned an

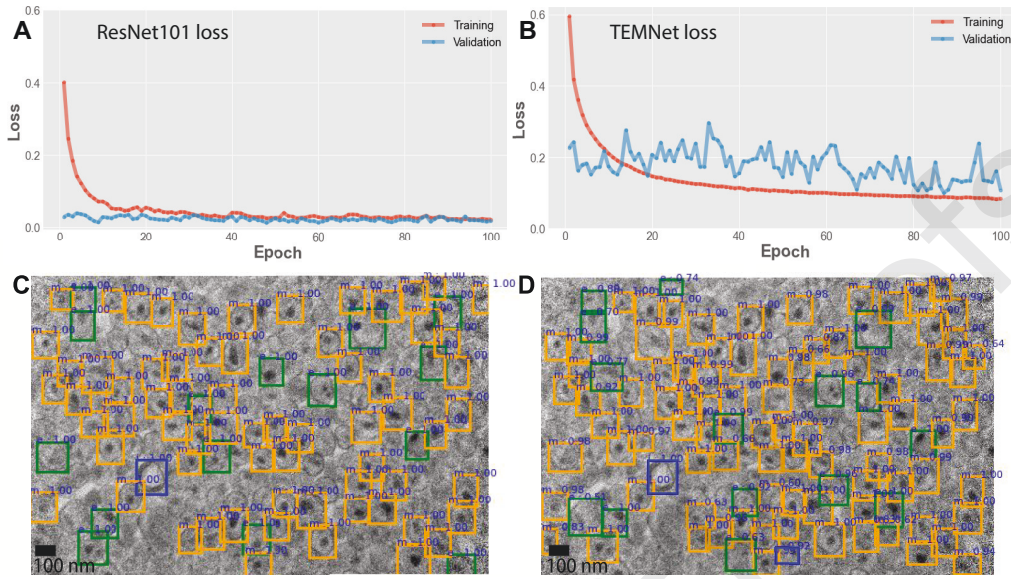


Figure 3: Learning error analysis of our particle classifier. **A** Training and validation error for the ResNet101 backbone versus training epoch, one full cycle of our Faster RCNN implementation. **B** Training and validation error for the TEMNet backbone versus training epoch. Early stopping and learning rate reduction were used to select the best weights avoiding overfitting. **C** Ground Truth labeled micrograph. All micrographs used for training our network were evaluated by eye and manually labeled. Instance classification of 200 HIV-1 particles required approximately 30 minutes to complete. **D** Automated predictions obtained using the TEMNet backbone for the micrograph in C. The network generated predictions on 130 micrographs in 5 minutes on one GPU. Scale bars in panels C and D are shown in the lower left portions of the images. Bounding box colors in panels C and D identify viral classifications: mature (yellow), immature (blue), eccentric (green).

input micrograph by translating a sliding window across the image and generating overlapping segmented regions. The segmented regions were compiled into batches and used as input for the RCNN network, which generated RoI (rectangular bounding box) coordinates and classification probability predictions for each virion instance detected in the segmented regions (see Figure A.11 for details). Then, the predicted RoIs were shifted by the position of the sliding window and gathered on the input full scale micrograph. Since the sliding window generated overlapping segmented regions, the network predicts multiple times on the virion instances localized in overlapping regions, generating overlapping RoIs with different classification probabilities that describe the same virion. To glean final predictions, non-max suppress-

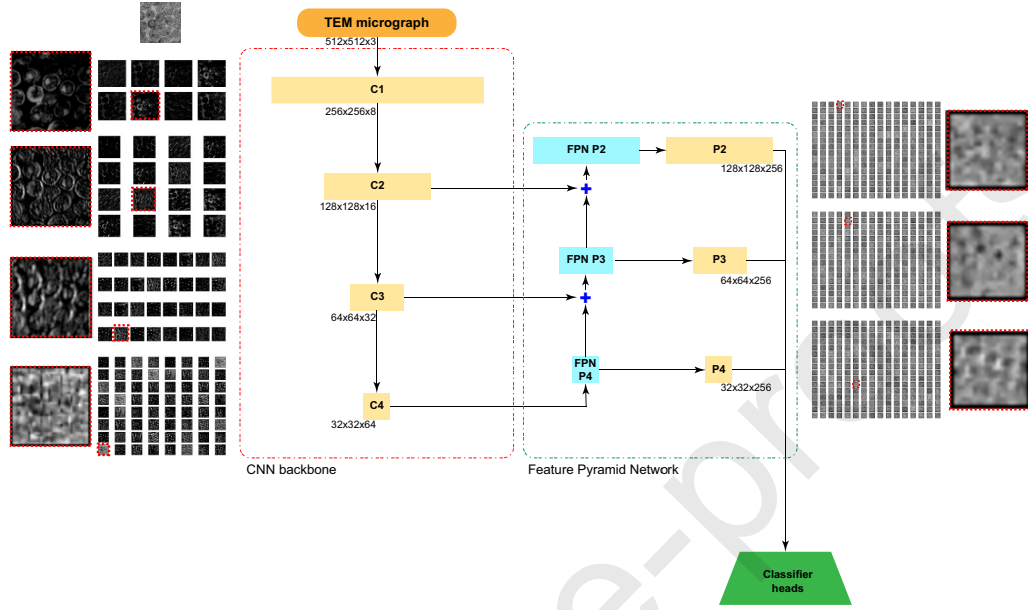


Figure 4: Feature map activations corresponding to each convolutional block of the TEM-Net backbone (C1, C2, C3, C4) and each level of the Feature Pyramid Network (P2, P3, P4). The first convolutional block identifies borders and electron density in virion lumina and isolates it from the background, the second block learns image processing techniques such as watershedding. Additional layers become more abstract and escape comprehensive description.

sion was applied to the predicted RoIs to eliminate the RoIs whose area overlapped more than a 30% threshold by retaining the RoI with the highest confidence score (i.e., prediction probability) and discarding the overlapping regions with lower confidence. In the case of confidence score ties, final RoIs were chosen by a larger area criterion due to a larger RoI being generally better at comprehending a viral instance in the full scale micrograph and providing better feature extraction through the Feature Pyramid Network. Finally, the resulting RoIs and class probability scores were displayed and a per class count was performed on the processed predictions. The default sliding window size for  $30,000\times$  magnification micrographs was  $1,024\times 1,024$  pixels ( $569\times 569$  nm) consistent with the cropping size and magnification used for building the training dataset.

The sliding window approach provided advantages to the prediction pipeline. For instance, translation variance of the predicted class for a given viral



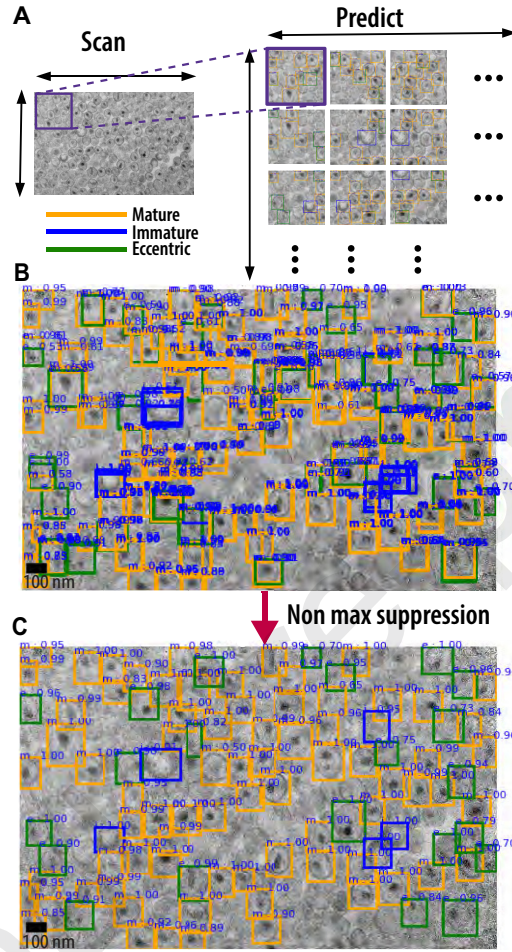


Figure 5: Micrograph segmentation via a sliding window: **A** A windowed region was translated across the image and predictions were generated on the segmented regions. **B** The predictions were gathered on the full scale micrograph and **C** Non-max suppression (NMS) was applied to determine classifications with highest confidence from overlapping Regions of Interest (RoIs), to glean final predictions. Numbers above each bounding box correspond to prediction "confidence" or certainty, which may ultimately be used to filter predictions (see Fig. 7). Scale bars are shown in the lower left portions of panels B and C.

instance was handled by considering predictions from different segmented regions and keeping the predicted RoI with the highest confidence score. Furthermore, this approach allowed for prediction generation on multi-magnification micrograph sets. Image size was linear with magnification, therefore a win-

dow size to magnification ratio  $r = W_t/M_t[px]$  can be calculated based on the cropping size  $W_t$  and magnification  $M_t$  used for training the network. This ratio was used internally by the network to calculate the appropriate sliding window size  $W_{new}$  for an input micrograph with a given magnification  $M_{new}$

$$W_{new} = rM_{new}[px], \quad (6)$$

the latter preserves the physical dimensions (in nm) of the segmented regions by the RCNN process, allowing consistent predictions across multiple magnifications. Samples of predicted micrographs at 30,000 $\times$  and 20,000 $\times$  magnification are presented in Figure 6. Additional predictions on 25,000 $\times$  magnification micrographs are presented in supplemental Figure A.12.

### 3.2. Prediction performance on raw images

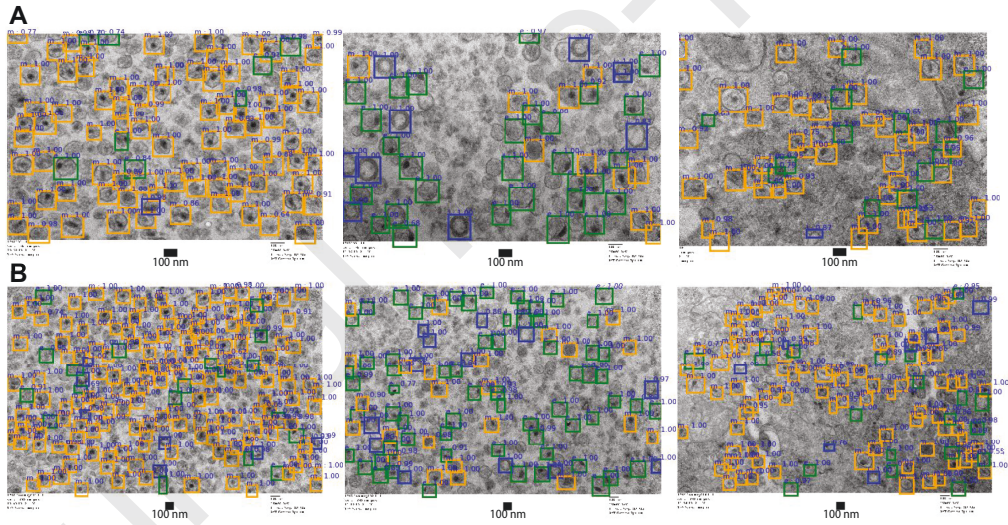


Figure 6: Virion classification on multi-magnification micrograph sets. Predictions across different raw TEM micrographs with **A** The same magnification used for training (30,000 $\times$ ). **B** A magnification lower than is discernible by a trained expert (20,000 $\times$ ). Our RCNN network calculates the appropriate sliding window size to segment a micrograph according to its magnification. Scale bars are shown beneath the micrographs.

To evaluate the performance of our network, we measured the mean Average Precision (mAP) as is traditional for object detection models[48]. mAP computes the average precision over all classes for a recall value from 0 to 1.

Precision, recall and F1 scores were calculated as

$$\begin{aligned}\text{Precision} &= \frac{\text{True Positives}}{\text{Total predicted instances}}, \\ \text{Recall} &= \frac{\text{True Positives}}{\text{Total ground truth instances}}, \\ \text{F1} &= 2 \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}\end{aligned}$$

where a *True Positive* prediction is established if its bounding box overlaps with a ground truth box by more than 50% Intersection over Union (IoU) [48] and the F1 score is the harmonic mean between precision and recall. Precision and recall pairs of values were calculated for increasing subsets of detections such that precision vs recall curves could be built from pair plotting. The precision vs recall curve was interpolated so dips in precision were replaced by the maximum precision for a given recall value. The mAP was then calculated as the area under the precision ( $p$ ) vs recall ( $r$ ) curve

$$mAP = \int_0^1 p_{\text{interpolated}}(r) dr, \quad (7)$$

since mAP takes into account precision, recall and the IoU overlapping of predicted RoIs to ground truth, it is regarded as the de-facto 'gold standard' to evaluate accuracy on Object Detection tasks across many datasets [49, 50, 51].

mAP for each of our Convolutional Neural Network backbones were measured on a validation dataset composed of 13 full scale TEM micrographs pertaining to different experimental conditions. As presented in Table 2, our model, TEMNet with a Batch Normalization (BN) layer in the Convolutional blocks achieved a 77.8% mAP competitive with InceptionResNetv2 (78.6% mAP), ResNet101 (78.2% mAP) and ResNet101v2 (77.9% mAP) while TEMNet with a Group Normalization (GN) layer outperformed these by over 1 point, achieving a 80.0% mAP. Regarding the other metrics ResNets achieved higher precision and F1 scores (for ResNet101 and ResNet101v2) while TEMNet achieved the highest recall score. Interestingly the InceptionResNetv2 architecture which was the most complex achieved the highest mAP among the ResNet backbones with also the highest recall. However, Inception-ResNet-v2 lacked the precision of other ResNet modules, perhaps due to the fact that zero padding was necessary to make the shapes of the inception blocks compatible with the upscaling and adding layers present in the FPN.



| RCNN backbone           | mAP <sub>IoU=0.5</sub> % | Precision % | Recall % | F1 score % |
|-------------------------|--------------------------|-------------|----------|------------|
| ResNet101 + FPN         | 78.2                     | 76.7        | 84.0     | 80.0       |
| ResNet101v2 + FPN       | 77.9                     | 75.8        | 85.9     | 80.5       |
| InceptionResNetv2 + FPN | 78.6                     | 64.7        | 86.2     | 73.7       |
| TEMNet (BN)             | 77.8                     | 60.8        | 84.9     | 70.9       |
| TEMNet (GN)             | 80.0                     | 71.3        | 87.9     | 78.8       |

Table 2: Object detection metrics for different Region based Convolutional Neural Network (RCNN) backbone architectures. The results represent the mean Average Precision (mAP) for predictions matching ground truth with an Intersection over Union (IoU) score over 0.5 tested on our validation dataset, the latter consisting of micrographs from different experimental conditions. Our model, TEMNet with a Batch Normalization (BN) in the Convolution blocks, achieved a mAP competitive with ResNet101 and ResNet101v2 while TEMNet with a Group Normalization (GN) layer outperformed ResNet101, ResNet101v2 and Inception-ResNet-v2 by over 1 mAP point.

In addition to being accurate, the network prediction pipeline derived in the present work significantly improved workflow times. While 30 minutes on average was required to manually ascribe 200 HIV-1 particle classifications on a single micrograph, our network offered a significant speedup, processing 130 micrographs in five minutes on one GPU, generating bounding box coordinates and classification probabilities as well as count histograms for each of the micrographs processed.

Furthermore, in order to evaluate the statistical distribution of classifications predicted by our Faster RCNN+FPN implementation, we measured the percentage of each particle morphology class across different experimental conditions. These morphology distributions were compared between the *in situ* ground truth classification counts from manually tabulated micrographs compared to the predicted distributions from the virion detection counts performed by our AI. Side by side histograms are shown in Figure 7 for predictions using the TEMNet backbone filtering predictions whose confidence score was above a 0.5 threshold. Ground truth labeled micrograph samples as well as their detection and classification predictions for each experiment are also presented in supplemental Figure A.13. Additionally, we compared the morphology distributions for an independent set of NL4-3 viruses and primary isolate samples that were not used as part of the training or validation micrographs. Side by side histograms are shown in Figure 8 for the TEMNet backbone with a confidence threshold of 0.5; histograms for other backbones with increasing confidence scores are presented in supplemental

Figure A.14.

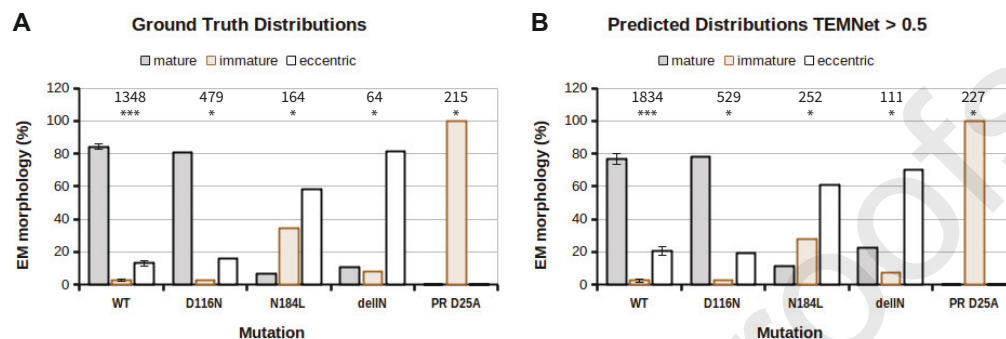


Figure 7: *In situ* classifications of virions from different HIV-1 IN (D116N, N184L, delIN) and PR mutant (D25A) viruses. **A** Ground truth distribution from manually ascribed micrograph sets. **B** Resulting distributions from TEMNet's predictions on the same micrographs. Predictions with a confidence score  $c$  above 0.5 were counted while those under this confidence threshold were rejected. Numbers over each distribution indicate the number of virus particles counted and the number of independent micrographs analyzed (\*). Error bars represent standard deviation from experimental replicates.

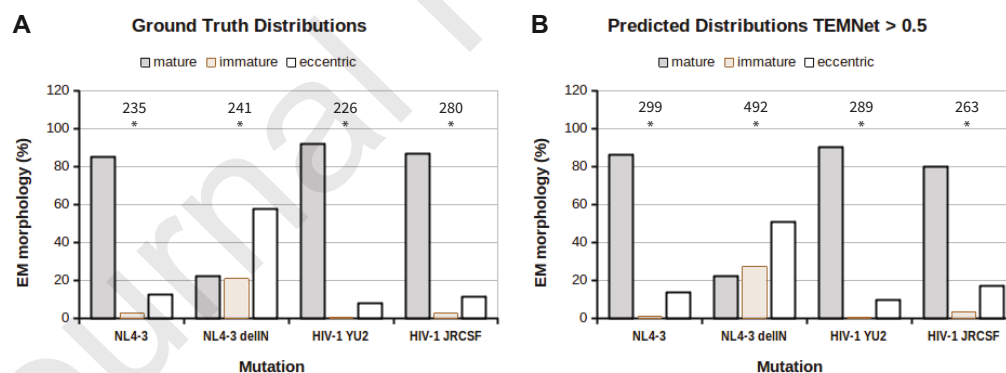


Figure 8: *In situ* classifications of virions from WT HIV-1<sub>NL4-3</sub>, IN deletion mutant delIN, and primary HIV-1 isolates YU2, and JR-CSF. **A** Ground truth distribution from manually ascribed micrograph sets. **B** Resulting distributions from TEMNet's predictions on the same micrographs. Predictions with a confidence score  $c$  above 0.5 were counted while those under this confidence threshold were rejected. Other indicators are the same as Figure 7.

|                     |            | RMSE (counts) |       |       |       |         |
|---------------------|------------|---------------|-------|-------|-------|---------|
| Backbone            | Confidence | WT            | D116N | N184L | delIN | PR D25A |
| TEMNet              | >0.5       | 7.75          | 2.70  | 4.73  | 9.69  | 0.00    |
|                     | >0.9       | 6.14          | 3.54  | 5.99  | 7.11  | 0.00    |
| ResNet101           | >0.5       | 1.42          | 1.09  | 3.73  | 5.11  | 0.00    |
|                     | >0.9       | 0.48          | 2.43  | 4.21  | 4.15  | 0.00    |
| ResNet101v2         | >0.5       | 3.51          | 1.60  | 1.73  | 2.71  | 0.00    |
|                     | >0.9       | 4.70          | 2.30  | 1.20  | 2.94  | 0.00    |
| Inception-ResNet-v2 | >0.5       | 10.28         | 2.49  | 7.54  | 3.32  | 0.00    |
|                     | >0.9       | 5.38          | 0.75  | 5.59  | 1.24  | 0.00    |

Table 3: Root Mean Square Error (RMSE) calculated between the predicted and ground truth distributions for each mutant virus. All predicted distributions were in accordance with the ground truth showing an error lower than 10% for all experiments independent of the convolutional backbone used. Among the convolutional backbones, ResNet101v2 provided the least error across all mutant viruses, followed by ResNet101, Inception-ResNet-v2 and TEMNet. Increasing the confidence threshold for which generated predictions were counted as True Positives for the distributions reduced the average RMSE across mutants for the TEMNet, ResNet101 and Inception-ResNet-v2 backbones and increased it for the ResNet101v2 backbone, helping in the first two cases to reduce the error for the WT and delIN mutants, which proved to be the most challenging micrographs to predict while PR D25A distributions were perfectly predicted (no error) due to the homogeneity of immature virions across these samples.

| Backbone            | Confidence | RMSE (counts) |             |      |        |
|---------------------|------------|---------------|-------------|------|--------|
|                     |            | NL4-3         | NL4-3 delIN | YU2  | JR-CSF |
| TEMNet              | >0.5       | 1.33          | 5.37        | 1.09 | 4.99   |
|                     | >0.9       | 3.18          | 5.96        | 2.72 | 2.14   |
| ResNet101           | >0.5       | 2.51          | 4.94        | 3.12 | 1.45   |
|                     | >0.9       | 2.77          | 4.42        | 4.04 | 3.40   |
| ResNet101v2         | >0.5       | 1.73          | 2.70        | 3.77 | 2.86   |
|                     | >0.9       | 1.92          | 2.20        | 4.11 | 2.62   |
| Inception-ResNet-v2 | >0.5       | 3.24          | 7.43        | 0.32 | 5.25   |
|                     | >0.9       | 1.21          | 5.53        | 3.39 | 2.17   |

Table 4: Root Mean Square Error (RMSE) calculated between the predicted and ground truth distributions for wild type HIV-1<sub>NL4-3</sub> and its IN deletion mutant delIN, as well as primary HIV-1 isolates YU2 and JR-CSF. Predicted distributions were in accordance with the ground truth distributions showing an error lower than 10% for all experiments independent of the convolutional backbone used. Among the convolutional backbones, ResNet101v2 provided the least error under these conditions for the HIV-1<sub>NL4-3</sub> samples, followed by ResNet101, TEMNet and Inception-ResNet-v2. At the same time, Inception-ResNet-v2 performed best with the primary virus samples closely followed by TEMNet. By increasing the confidence score threshold, TEMNet performed best for HIV-1<sub>NL4-3</sub> and HIV-1<sub>YU2</sub> samples, while ResNet101 performed best with HIV-1<sub>JR-CSF</sub> samples.

As summarized in Tables 3 and 4, ascribed morphology distributions were in accordance with the associated ground truth measures with a root mean squared error (RMSE) lower than 10% for each virus type independent of the CNN backbone used in the Faster RCNN architecture. Among the analyzed viruses, WT NL4-3 and delIN were the most challenging to classify, which corresponded to micrographs that presented the lowest contrast between particle instances and background along with the most image noise out of the validation samples (see Figure A.13). By contrast, PR D25A mutant viral samples, which consisted of only immature virions, were classified without error. For the TEMNet, ResNet101 and Inception-ResNet-v2 backbones, the error for WT and delIN samples could be reduced by increasing the confidence threshold for which True Positive instances were counted, reducing the average RMSE across all samples for these backbones but increasing it for the ResNet101v2 backbone. Similarly, increasing the confidence threshold reduces the average RMSE for the primary isolates using the TEMNet backbone. In this regard TEMNet proves especially important providing the lowest error when predicting the morphologies of the primary isolate viral samples. Additionally, the calculated  $p$ -value for the Pearson’s  $\chi^2$  test (see Tables 5 and 6) indicates no significant statistical difference between the ground truth and predicted distributions across the different mutant viruses, with the exception of the WT virus when using the TEMNet backbone. In this regard, ResNet101v2 proved to be the most accurate backbone when comparing the predicted morphology distributions across different experimental conditions, followed by ResNet101, Inception-ResNet-v2 and TEMNet. For this reason, we consider that ResNets outperform TEMNet on prediction accuracy. However, while the former algorithms generally predicted virion morphology more accurately across samples, TEMNet importantly offered an accuracy trade off for prediction speed and compactivity.

Histograms for all backbones across different experimental samples with varying confidence score threshold are presented in supplemental Figures A.15 and A.16.

#### 4. Discussion

We have developed an end-to-end deep learning solution to the automated detection and classification of HIV-1 virion particle morphologies from TEM micrographs across different maturation stages. Our overall methodology is not limited to HIV-1 particles and can be extended to other enveloped viruses

| Backbone            | Confidence | $\chi^2$ test $p$ -value |       |       |       |         |
|---------------------|------------|--------------------------|-------|-------|-------|---------|
|                     |            | WT                       | D116N | N184L | delIN | PR D25A |
| TEMNet              | >0.5       | 0.02                     | 0.34  | 0.07  | 0.08  | 1.00    |
|                     | >0.9       | 0.04                     | 0.21  | 0.06  | 0.12  | 1.00    |
| ResNet101           | >0.5       | 0.70                     | 0.91  | 0.15  | 0.43  | 1.00    |
|                     | >0.9       | 0.92                     | 0.56  | 0.17  | 0.56  | 1.00    |
| ResNet101v2         | >0.5       | 0.27                     | 0.80  | 0.67  | 0.77  | 1.00    |
|                     | >0.9       | 0.08                     | 0.63  | 0.84  | 0.73  | 1.00    |
| Inception-ResNet-v2 | >0.5       | $6.33\text{e}^{-4}$      | 0.64  | 0.01  | 0.66  | 1.00    |
|                     | >0.9       | $6.50\text{e}^{-3}$      | 0.89  | 0.08  | 0.90  | 1.00    |

Table 5: Pearson’s  $\chi^2$  test  $p$ -values calculated between the predicted and ground truth distributions for each experimental condition. Within the probability threshold of  $p = 5\%$  there was no statistical difference between the distributions calculated from the predictions and the distributions calculated from ground truth counts with the exception of the WT virus with the TEMNet backbone. In accordance with the RMSE values (Table 3) for each backbone, ResNet101v2 provided the highest  $p$ -value followed by ResNet101, Inception-ResNet-v2 and TEMNet.  $p$ -values greater than 0.05 indicate no significant statistical difference.

| Backbone            | Confidence | $\chi^2$ test $p$ -value |                     |                     |                     |
|---------------------|------------|--------------------------|---------------------|---------------------|---------------------|
|                     |            | NL4-3                    | NL4-3 delIN         | YU2                 | JR-CSF              |
| TEMNet              | >0.5       | 0.18                     | 0.03                | 0.76                | 0.02                |
|                     | >0.9       | 0.08                     | 0.01                | 0.06                | 0.28                |
| ResNet101           | >0.5       | 0.29                     | 0.02                | 0.02                | 0.26                |
|                     | >0.9       | 0.21                     | 0.05                | $2.03\text{e}^{-4}$ | 0.02                |
| ResNet101v2         | >0.5       | 0.58                     | 0.46                | $5.37\text{e}^{-4}$ | $1.30\text{e}^{-4}$ |
|                     | >0.9       | 0.51                     | 0.54                | $3.69\text{e}^{-5}$ | $2.78\text{e}^{-3}$ |
| Inception-ResNet-v2 | >0.5       | 0.09                     | $1.90\text{e}^{-5}$ | $1.32\text{e}^{-6}$ | 0.02                |
|                     | >0.9       | 0.02                     | $2.57\text{e}^{-3}$ | $2.91\text{e}^{-6}$ | 0.44                |

Table 6: Pearson’s  $\chi^2$  test  $p$ -values calculated between the predicted and ground truth distributions for wild type NL4-3 and primary isolate YU2 and JR-CSF HIV-1 viruses. TEMNet is the only backbone that predicts a distribution with no statistical significance for primary isolates within a probability threshold of  $p = 5\%$  when increasing the confidence threshold to 0.9. The highest average  $p$ -values are encountered for ResNet101v2, TEMNet, ResNet101 and Inception-ResNet-v2 in descending order.  $p$ -values greater than 0.05 indicate no significant statistical difference.

provided that enough training data is available. In our approach, we have overcome the limitations of comparatively small datasets to produce reliable particle classifications and counts.

Our network, named TEMNet, is a new CNN architecture for object detection and has been trained from scratch as a backbone for a two-stage Faster RCNN [29] object detection network. In line with [37] and [38], we demonstrated that our model converges when trained from scratch thanks to Group Normalization [46] techniques and building a reasonably sized dataset consisting of 13,650 labeled croppings of TEM micrographs for training and validation. Importantly, the training dataset was built from different experiments. Outcomes of particle classification were pitted head-to-head versus manual ascriptions of the same micrographs, allowing the model to be robust and generalizable for HIV-1 virions under diverse experimental setups.

We have demonstrated that networks developed to handle photon-based images are competent at identifying and classifying objects from electron-based imaging. Comparing TEMNet with ImageNet pretrained ResNet [42, 43, 44] backbones we found that while both networks worked with a high accuracy on validation micrographs from different experiments, TEMNet reported the highest mAP score at 80.0% surpassing ResNets by over 1 mAP point (Table 2). All backbones predicted statistically significant data when comparing the predicted morphology percentages for in situ micrographs for different IN and PR mutant viruses, with manually ascribed ground truth distributions. In this regard, however, ResNet backbones outperformed TEMNet, presenting the lowest RMSE and the highest  $p$ -values for the Pearson  $\chi^2$  test as summarized in Tables 3 and 5. The WT and delIN mutant viruses, whose samples had the most noise, performed the poorest across techniques, while the PR D25A active site mutant virus performed best, owing to the uniformity of the immature particle morphology across samples (Figures 7, A.15 and A.16).

Faster RCNN combined with our TEMNet backbone also proved to be a highly efficient method for generating predictions on raw TEM micrographs, offering a significant speedup to manual classification. While 30 minutes on average was required to manually ascribe each micrograph that contained approximately 200 viral particles, the model processed and generated predictions of 130 micrographs in five minutes with a single GPU. This translates to an approximate 780-fold improvement in speed. In addition, our prediction method could also handle particle predictions for multi-magnification micrograph sets, demonstrated in Figures 6 and A.12. Finally, the TEMNet

backbone was accurate, efficient and also light. The memory footprint of TEMNet’s training weights was only 15 MB compared to ResNet’s 235 MB and Inception-ResNet-v2’s 292 MB, which renders TEMNet appropriate for software implementations under hardware constraints and therefore useful for web and mobile deployment.

Summarizing, here we present a robust Convolutional Neural Network for the automated detection and classification of HIV-1 particle morphologies from TEM micrographs. Our proposed TEMNet backbone has the capability to accurately and efficiently detect HIV-1 virions and classify them according to their maturation stage across varying experimental conditions. Furthermore, the statistical distributions across experimental conditions agreed with manually ascribed results while being significantly faster. Given that Gag-interacting maturation inhibitors and ALLINIs, each of which disrupt particle maturation, are in preclinical development, our methodology could prove useful in highly promising antiretroviral drug development programs. We moreover expect that our tool could prove useful to a broader range of scientists including virologist and medical researchers, as long as there is sufficient raw data on which to first train the machine learning methodology. The latter could especially apply to histopathological detection of SARS-CoV-2 infection (see [52] for review), where cell organelles that are similar in size to virus particles often confound data interpretation.

## 5. Acknowledgments

The authors acknowledge funding from the US National Institutes of Health awards R01AI070042 (to A.N.E.), P50AI1504817 (to A.N.E. and J.R.P.) and P20GM104316 (to J.R.P.). Funding for methodology development for the present work was provided by the National Science Foundation award MCB-2027096, funded in part by Delaware Established Program to Stimulate Competitive Research (EPSCoR). This research is part of the Frontera computing project at the Texas Advanced Computing Center. Frontera is made possible by NSF award OAC-1818253. This work used the Extreme Science and Engineering Discovery Environment, which is supported by the National Science Foundation (Grant ACI-1548562). Christian Lantz was supported by NSF award CHEM-1560325. The present work was also funded by the Gordon and Betty Moore Foundation and the Research Corporation. The authors also thank Gulcin Pekurnnaz and Abhishek Chatterjee for insightful discussions.



## References

- [1] F. P. O. Nagler, G. Rake, The use of the electron microscope in diagnosis of variola, vaccinia, and varicella, *Journal of Bacteriology* 55 (1) (1948) 45–51.
- [2] S. Brenner, R. Horne, A negative staining method for high resolution electron microscopy of viruses, *Biochimica et Biophysica Acta* 34 (1959) 103–110. doi:[https://doi.org/10.1016/0006-3002\(59\)90237-9](https://doi.org/10.1016/0006-3002(59)90237-9).
- [3] F. Barre-Sinoussi, J. Chermann, F. Rey, M. Nugeyre, S. Chamaret, J. Gruest, C. Dautet, C. Axler-Blin, F. Vezinet-Brun, C. Rouzioux, W. Rozenbaum, L. Montagnier, Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (aids), *Science* 220 (4599) (1983) 868–871. doi:[10.1126/science.6189183](https://doi.org/10.1126/science.6189183).
- [4] S. von Stillfried, P. Boor, Detection methods for SARS-CoV-2 in tissue, *Der Pathologe* (2021) 1432–1963doi:[10.1007/s00292-021-00920-1](https://doi.org/10.1007/s00292-021-00920-1).
- [5] W. Sundquist, H.-G. Kräusslich, HIV-1 assembly, budding, and maturation, *Cold Spring Harbor perspectives in medicine* 2 (2012) a006924. doi:[10.1101/cshperspect.a006924](https://doi.org/10.1101/cshperspect.a006924).
- [6] M. A. Accola, Å. Öhagen, H. G. Göttinger, Isolation of human immunodeficiency virus type 1 cores: Retention of vpr in the absence of p6gag, *Journal of Virology* 74 (13) (2000) 6198–6202. doi:[10.1128/JVI.74.13.6198-6202.2000](https://doi.org/10.1128/JVI.74.13.6198-6202.2000).
- [7] J. Fontana, K. A. Jurado, N. Cheng, N. L. Ly, J. R. Fuchs, R. J. Gorelick, A. N. Engelman, A. C. Steven, Distribution and redistribution of HIV-1 nucleocapsid protein in immature, mature, and integrase-inhibited virions: a role for integrase in maturation, *Journal of Virology* 89 (19) (2015) 9765–9780. doi:[10.1128/JVI.01522-15](https://doi.org/10.1128/JVI.01522-15).
- [8] J. Anderson, C. Schiffer, S.-K. Lee, R. Swanstrom, *Viral Protease Inhibitors*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2009, pp. 85–110. doi:[10.1007/978-3-540-79086-0\\_4](https://doi.org/10.1007/978-3-540-79086-0_4).
- [9] A. B. Kleinpeter, E. O. Freed, HIV-1 maturation: Lessons learned from inhibitors, *Viruses* 12 (9) (2020). doi:[10.3390/v12090940](https://doi.org/10.3390/v12090940).

- [10] A. Engelman, G. Englund, J. M. Orenstein, M. A. Martin, R. Craigie, Multiple effects of mutations in human immunodeficiency virus type 1 integrase on viral replication., *Journal of Virology* 69 (5) (1995) 2729–2736.
- [11] K. A. Jurado, H. Wang, A. Slaughter, L. Feng, J. J. Kessl, Y. Koh, W. Wang, A. Ballandras-Colas, P. A. Patel, J. R. Fuchs, M. Kvaratskhelia, A. Engelman, Allosteric integrase inhibitor potency is determined through the inhibition of HIV-1 particle maturation, *Proceedings of the National Academy of Sciences* 110 (21) (2013) 8690–8695. doi:10.1073/pnas.1300703110.
- [12] M. Balakrishnan, S. R. Yant, L. Tsai, C. O’Sullivan, R. A. Bam, A. Tsai, A. Niedziela-Majka, K. M. Stray, R. Sakowicz, T. Cihlar, Non-catalytic site HIV-1 integrase inhibitors disrupt core maturation and induce a reverse transcription block in target cells, *PLoS one* 8 (9) (2013) 1–12.
- [13] J. L. Elliott, J. E. Eschbach, P. C. Koneru, W. Li, M. Puray-Chavez, D. Townsend, D. Q. Lawson, A. N. Engelman, M. Kvaratskhelia, S. B. Kutluay, Integrase-RNA interactions underscore the critical role of integrase in HIV-1 virion morphogenesis, *eLife* 9 (September 2020). doi:10.7554/elife.54311.
- [14] J. J. Kessl, S. B. Kutluay, D. Townsend, S. Rebersburg, A. Slaughter, R. C. Larue, N. Shkriabai, N. Bakouche, J. R. Fuchs, P. D. Bieniasz, M. Kvaratskhelia, HIV-1 integrase binds the viral RNA genome and is essential during virion morphogenesis, *Cell* 166 (5) (2016) 1257–1268.e12. doi:10.1016/j.cell.2016.07.044.
- [15] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, C. I. Sánchez, A survey on deep learning in medical image analysis, *Medical Image Analysis* 42 (2017) 60–88. doi:https://doi.org/10.1016/j.media.2017.07.005.
- [16] M. Liu, F. Li, H. Yan, K. Wang, Y. Ma, L. Shen, M. Xu, A. D. N. Initiative, et al., A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in Alzheimer’s disease, *NeuroImage* 208 (2020) 116459.

- [17] G. Jain, D. Mittal, D. Thakur, M. K. Mittal, A deep learning approach to detect COVID-19 coronavirus with X-ray images, *Biocybernetics and Biomedical Engineering* 40 (4) (2020) 1391–1405. doi:<https://doi.org/10.1016/j.bbe.2020.08.008>.
- [18] M. Veta, P. J. van Diest, R. Kornegoor, A. Huisman, M. A. Viergever, J. P. W. Pluim, Automatic nuclei segmentation in h&e stained breast cancer histopathology images, *PLoS One* 8 (7) (07 2013). doi:[10.1371/journal.pone.0070221](https://doi.org/10.1371/journal.pone.0070221).
- [19] F. Xing, Y. Xie, L. Yang, An automatic learning-based framework for robust nucleus segmentation, *IEEE transactions on medical imaging* 35 (09 2015). doi:[10.1109/TMI.2015.2481436](https://doi.org/10.1109/TMI.2015.2481436).
- [20] E. Ito, T. Sato, D. Sano, E. Utagawa, T. Kato, Virus particle detection by convolutional neural network in transmission electron microscopy images, *Food and Environmental Virology* 10 (2018) 1–8. doi:[10.1007/s12560-018-9335-7](https://doi.org/10.1007/s12560-018-9335-7).
- [21] L. Nanni, E. De Luca, M. L. Facin, G. Maguolo, Deep learning and handcrafted features for virus image classification, *Journal of Imaging* 6 (12) (2020) 143.
- [22] E. Gómez de Mariscal, M. Maška, A. Kotrbová, V. Pospichalova, P. Matula, A. Muñoz-Barrutia, Deep-learning-based segmentation of small extracellular vesicles in transmission electron microscopy images, *Scientific Reports* 9 (2019) 1–10. doi:[10.1038/s41598-019-49431-3](https://doi.org/10.1038/s41598-019-49431-3).
- [23] I. Nikishin, R. Dulimov, G. Skryabin, S. Galetsky, E. Tchevkina, D. Bagrov, ScanEV – a neural network-based tool for the automated detection of extracellular vesicles in tem images, *Micron* 145 (2021) 103044. doi:<https://doi.org/10.1016/j.micron.2021.103044>.
- [24] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: N. Navab, J. Hornegger, W. M. Wells, A. F. Frangi (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Springer International Publishing, Cham, 2015, pp. 234–241.

- [25] J. Xu, L. Xiang, R. Hang, J. Wu, Stacked sparse autoencoder (SSAE) based framework for nuclei patch classification on breast cancer histopathology, in: 2014 IEEE 11th International Symposium on Biomedical Imaging, ISBI 2014, 2014, pp. 999–1002. doi:10.1109/ISBI.2014.6868041.
- [26] Y. Xie, Z. Zhang, M. Sapkota, L. Yang, Spatial clockwork recurrent neural network for muscle perimysium segmentation, in: S. Ourselin, L. Joskowicz, M. R. Sabuncu, G. Unal, W. Wells (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016, Springer International Publishing, Cham, 2016, pp. 185–193.
- [27] F. Xing, Y. Xie, H. Su, F. Liu, L. Yang, Deep learning in microscopy image analysis: A survey, *IEEE Transactions on Neural Networks and Learning Systems* 29 (10) (2018) 4550–4568. doi:10.1109/TNNLS.2017.2766168.
- [28] I. Moreno-Indias, L. Lahti, M. Nedyalkova, I. Elbere, G. Roshchupkin, M. Adilovic, O. Aydemir, B. Bakir-Gungor, E. C.-d. Santa Pau, D. D’Elia, M. S. Desai, L. Falquet, A. Gundogdu, K. Hron, T. Klammersteiner, M. B. Lopes, L. J. Marcos-Zambrano, C. Marques, M. Mason, P. May, L. Pašić, G. Pio, S. Pongor, V. J. Promponas, P. Przyklus, J. Saez-Rodriguez, A. Sampri, R. Shigdel, B. Stres, R. Suharoschi, J. Truu, C.-O. Truică, B. Vilne, D. Vlachakis, E. Yilmaz, G. Zeller, A. L. Zomer, D. Gómez-Cabrero, M. J. Claesson, Statistical and machine learning techniques in human microbiome studies: Contemporary challenges and solutions, *Frontiers in Microbiology* 12 (2021) 277. doi:10.3389/fmicb.2021.635781.
- [29] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: Towards real-time object detection with region proposal networks, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39 (6) (2017) 1137–1149. doi:10.1109/TPAMI.2016.2577031.
- [30] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, et al., Tensorflow: A system for large-scale machine learning, in: 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16), 2016, pp. 265–283.

- [31] A. Adachi, H. Gendelman, S. Koenig, T. Folks, R. Willey, A. Rabson, M. Martin, Production of acquired immunodeficiency syndrome-associated retrovirus in human and nonhuman cells transfected with an infectious molecular clone, *Journal of Virology* 59 (2) (1986) 284–291. doi:10.1128/jvi.59.2.284-291.1986.
- [32] H. E. V. Brown, H. Chen, A. Engelman, Structure-based mutagenesis of the human immunodeficiency virus type 1 DNA attachment site: Effects on integration and cDNA synthesis, *Journal of Virology* 73 (11) (1999) 9011–9020. doi:10.1128/JVI.73.11.9011-9020.1999.
- [33] Y. Li, J. Kappes, J. Conway, R. Price, G. Shaw, B. Hahn, Molecular characterization of human immunodeficiency virus type 1 cloned directly from uncultured human brain tissue: identification of replication-competent and -defective viral genomes, *Journal of Virology* 65 (8) (1991) 3973–3985. doi:10.1128/jvi.65.8.3973-3985.1991.
- [34] Y. Koyanagi, S. Miles, R. Mitsuyasu, J. Merrill, H. Vinters, I. Chen, Dual infection of the central nervous system by AIDS viruses with distinct cellular tropisms, *Science (New York, N.Y.)* 236 (4803) (1987) 819–822. doi:10.1126/science.3646751.
- [35] J. Anderson-Daniels, P. K. Singh, G. A. Sowd, W. Li, A. N. Engelman, C. Aiken, Dominant negative MA-CA fusion protein is incorporated into HIV-1 cores and inhibits nuclear entry of viral preintegration complexes, *Journal of virology* 93 (21) (November 2019). doi:10.1128/jvi.01118-19.
- [36] J. Deng, W. Dong, R. Socher, L. Li, Kai Li, Li Fei-Fei, ImageNet: A large-scale hierarchical image database, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 248–255. doi:10.1109/CVPR.2009.5206848.
- [37] Z. Shen, Z. Liu, J. Li, Y. G. Jiang, Y. Chen, X. Xue, Object detection from scratch with deep supervision, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 42 (2) (2020) 398–412. doi:10.1109/TPAMI.2019.2922181.
- [38] K. He, R. Girshick, P. Dollar, Rethinking ImageNet pre-training, in: 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 4917–4926. doi:10.1109/ICCV.2019.00502.

- [39] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, in: 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 580–587. doi:10.1109/CVPR.2014.81.
- [40] R. Girshick, Fast R-CNN, in: 2015 IEEE International Conference on Computer Vision (ICCV), 2015, pp. 1440–1448. doi:10.1109/ICCV.2015.169.
- [41] K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask R-CNN, IEEE Transactions on Pattern Analysis and Machine Intelligence 42 (2) (2020) 386–397. doi:10.1109/TPAMI.2018.2844175.
- [42] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778. doi:10.1109/CVPR.2016.90.
- [43] K. He, X. Zhang, S. Ren, J. Sun, Identity mappings in deep residual networks, in: B. Leibe, J. Matas, N. Sebe, M. Welling (Eds.), Computer Vision – ECCV 2016, Springer International Publishing, Cham, 2016, pp. 630–645.
- [44] C. Szegedy, S. Ioffe, V. Vanhoucke, A. A. Alemi, Inception-v4, inception-resnet and the impact of residual connections on learning, in: Thirty-first AAAI conference on artificial intelligence, 2017.
- [45] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, in: F. Bach, D. Blei (Eds.), Proceedings of the 32nd International Conference on Machine Learning, Vol. 37 of Proceedings of Machine Learning Research, PMLR, Lille, France, 2015, pp. 448–456.
- [46] Y. Wu, K. He, Group normalization, International Journal of Computer Vision 128 (03 2020). doi:10.1007/s11263-019-01198-w.
- [47] T. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, S. Belongie, Feature pyramid networks for object detection, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 936–944. doi:10.1109/CVPR.2017.106.

- [48] M. Everingham, S. Eslami, L. Van Gool, C. Williams, J. Winn, A. Zisserman, The PASCAL visual object classes challenge: A retrospective, *International Journal of Computer Vision* 111 (1) (2015) 98–136. doi:10.1007/s11263-014-0733-5.
- [49] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C. L. Zitnick, Microsoft COCO: Common objects in context, in: D. Fleet, T. Pajdla, B. Schiele, T. Tuytelaars (Eds.), *Computer Vision – ECCV 2014*, Springer International Publishing, Cham, 2014, pp. 740–755.
- [50] M. Everingham, L. V. Gool, C. K. I. Williams, J. Winn, A. Zisserman, The Pascal visual object classes (VOC) challenge, *International Journal of Computer Vision* 88 (2009) 303–308, printed version publication date: June 2010.
- [51] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. Berg, L. Fei-Fei, ImageNet large scale visual recognition challenge, *International Journal of Computer Vision* 115 (09 2014). doi:10.1007/s11263-015-0816-y.
- [52] H. Hopfer, M. C. Herzig, R. Gosert, T. Menter, J. Hench, A. Tzankov, H. H. Hirsch, S. E. Miller, Hunting coronavirus by transmission electron microscopy - a guide to SARS-CoV-2-associated ultrastructural pathology in COVID-19 tissues, *Histopathology* 78 (3) (2021) 358–370. doi:10.1111/his.14264.



## Appendix A. Supplemental Figures

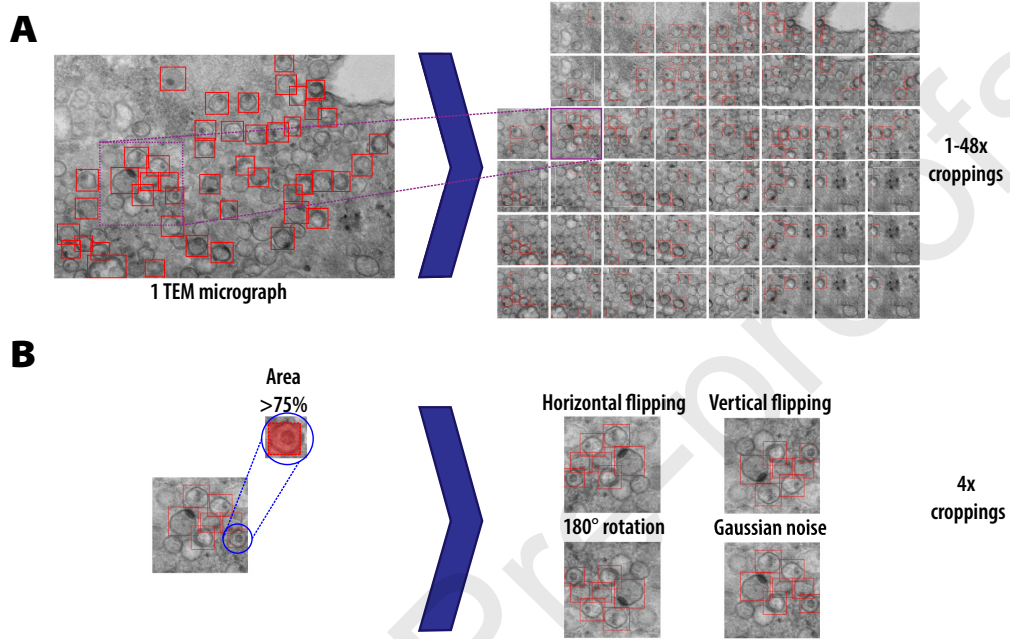


Figure A.9: Data augmentation procedure applied to TEM micrographs: **A** A raw TEM micrograph is cropped into overlapping regions of  $1024 \times 1024$  pixels containing viral particles, a viral particle is counted as ground truth inside a cropped region when at least 75% of its bounding box area is contained within the region. (1-48x augmentation). **B** Cropped images are further augmented by means of transformations: vertical flipping, horizontal flipping,  $180^\circ$  rotations and gaussian noise. (4x augmentation).



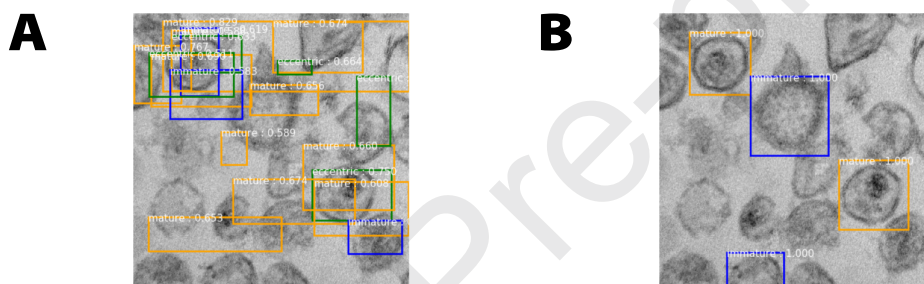


Figure A.10: Importance of pretraining for CNN architectures: **A** TEM micrograph cropping processed displaying predicted HIV-1 viral particles using ResNet101 weights trained from scratch on a small dataset. **B** The same TEM micrograph section processed displaying HIV-1 viral particles predicted by RCNN with a ResNet101 backbone pretrained on the massive-scale dataset ImageNet.

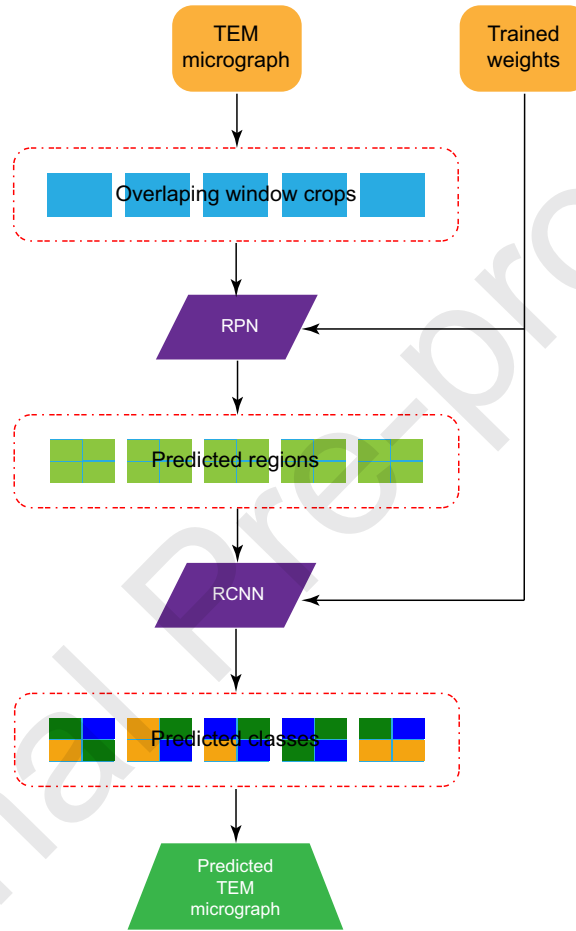


Figure A.11: Flowchart representing the prediction algorithm of the RCNN network: A TEM micrograph is divided into overlapping window crops that are passed to the RPN, which detects regions with viral instances; these are then passed to the RCNN classifier head which returns a class for each of the regions. Then predictions are aggregated and processed into a predicted TEM micrograph. All RPN and RCNN classifiers use trained weights to generate predictions.

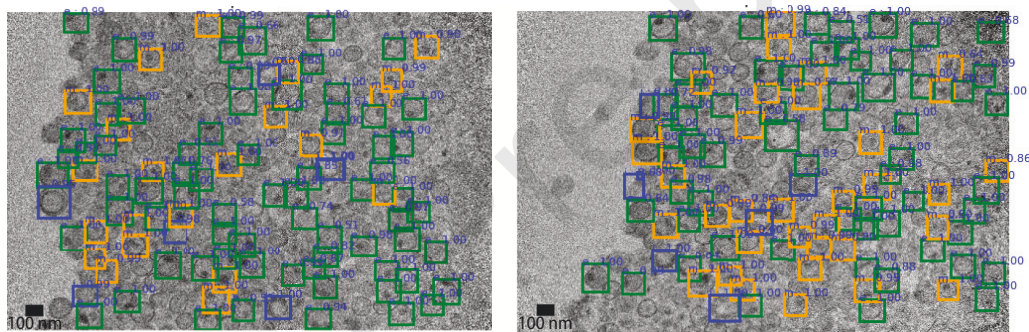


Figure A.12: Predicted classification of HIV-1 virions for TEM micrographs at 25000 $\times$  magnification using TEMNet.

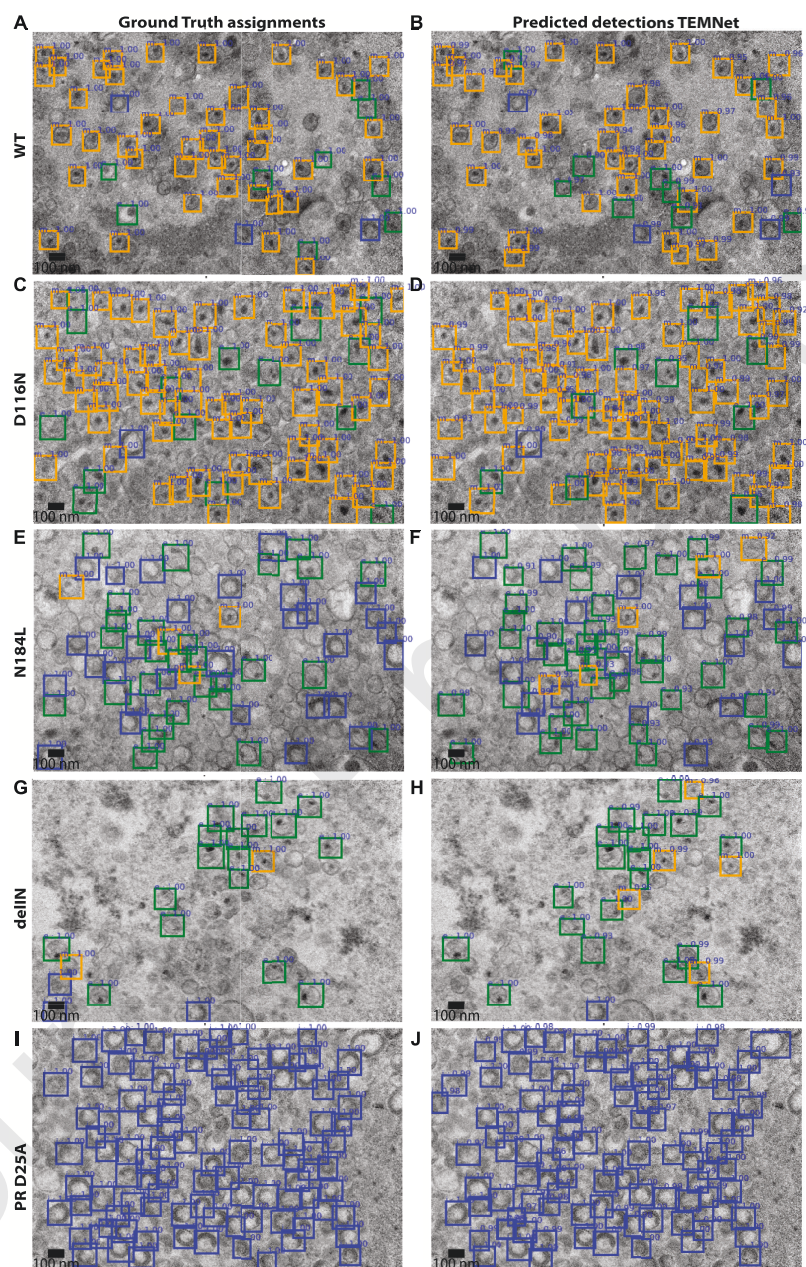


Figure A.13: *In situ* detection and classifications of virions from different WT and mutant HIV-1 samples. Ground truth labeled micrograph from an expertly trained eye for **A** WT, **C** D116N, **E** N184L, **G** delIN and **I** PR D25A viruses. Detection and classification predictions generated from TEMNet on the same **B** WT, **D** D116N, **F** N184L, **H** delIN and **J** PR D25A samples.

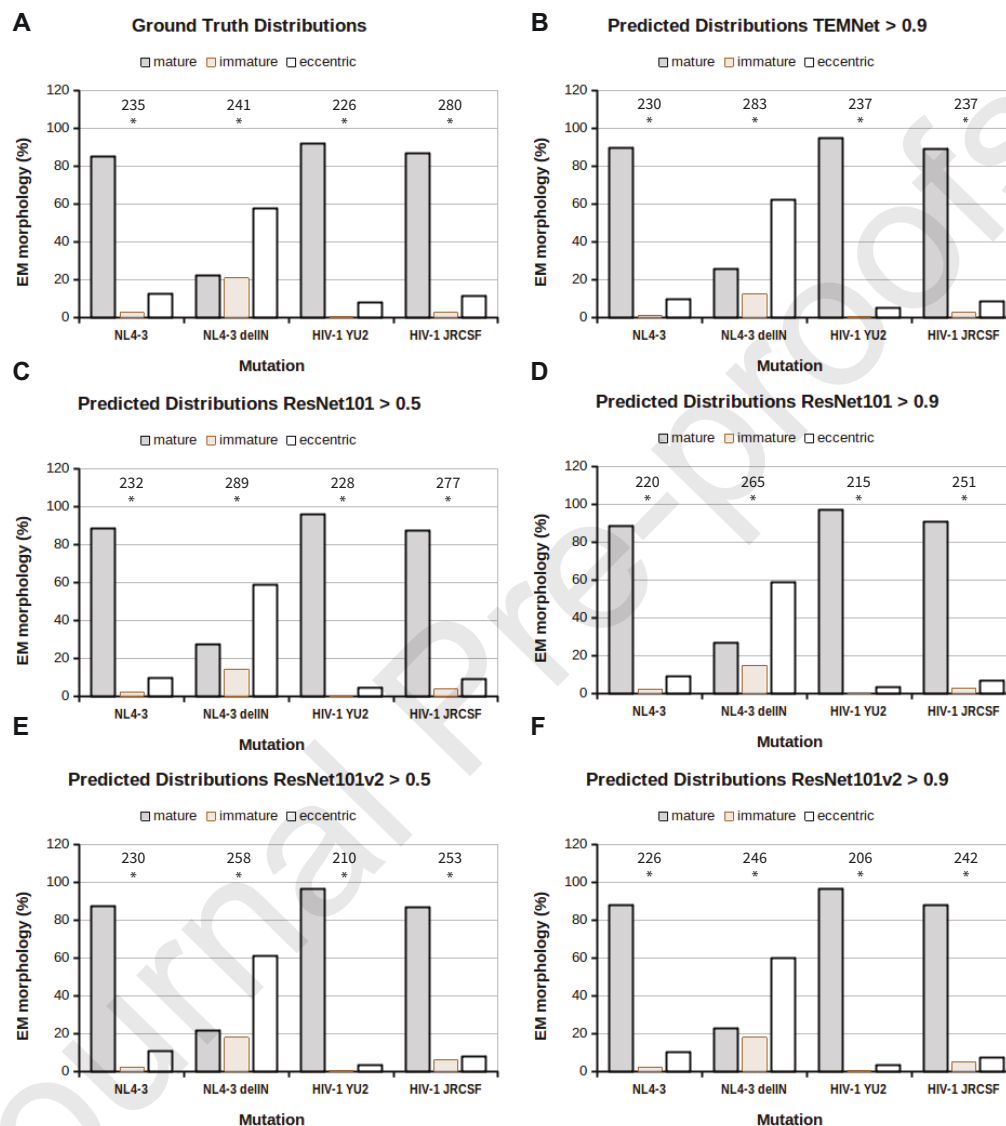


Figure A.14: *In situ* classifications of WT NL4-3 and NL4-3 delIN and primary isolates YU2 and JR-CSF HIV-1 virions. **A** Ground truth distribution from manually labeled micrographs. **B-F** Resulting distributions from different CNN backbones with a confidence score threshold  $c$ : **B** TEMNet,  $c > 0.9$ . **C** ResNet101,  $c > 0.5$ . **D** ResNet101,  $c > 0.9$ . **E** ResNet101v2,  $c > 0.5$ . **F** ResNet101v2,  $c > 0.9$ . Error bars, numbers and asterisks are as in Figure 7.

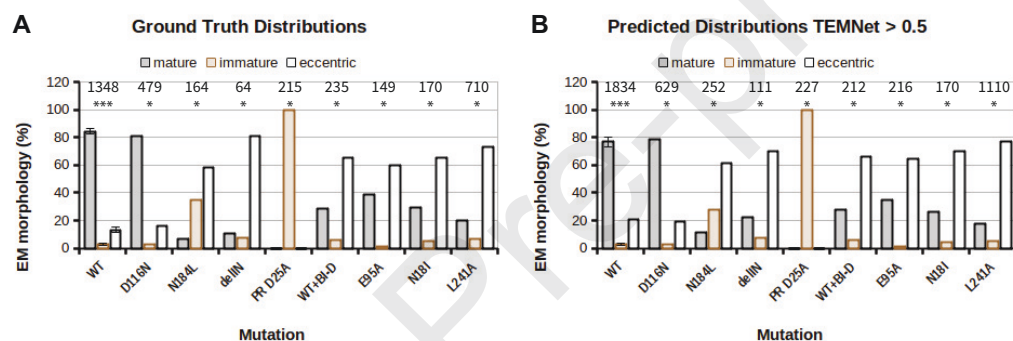


Figure A.15: *In situ* classifications of WT and IN (D116N, N184L, delIN, E95A, N18I, L241A), WT+BI-D and PR D25A mutant virions. **A** Ground truth distribution from manually labeled micrographs. **B** Resulting distributions from TEMNet's predictions on the same micrographs. Predictions with a confidence score  $c$  above 0.5 were counted while those under this confidence threshold were rejected. See Figure 7 for meanings of numbers, asterisks, and error bars.



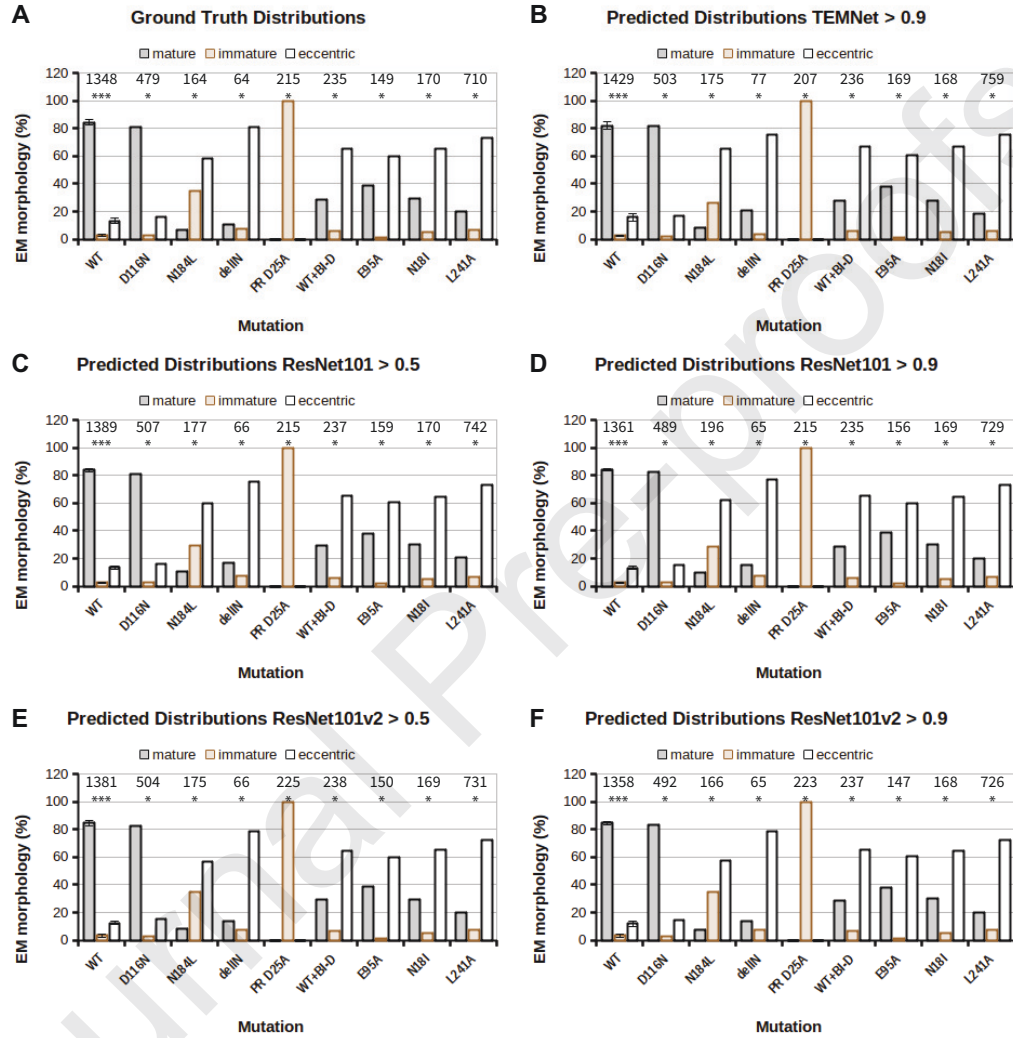


Figure A.16: *In situ* classifications of WT and IN (D116N, N184L, delIN, E95A, N181, L241A), WT+BI-D and PR D25A mutant virions. **A** Ground truth distribution from manually labeled micrographs. **A** Ground truth distribution from manually labeled micrographs. **B-F** Resulting distributions from different CNN backbones with a confidence score threshold  $c$ : **B** TEMNet,  $c > 0.9$ . **C** ResNet101,  $c > 0.5$ . **D** ResNet101,  $c > 0.9$ . **E** ResNet101v2,  $c > 0.5$ . **F** ResNet101v2,  $c > 0.9$ . See Figure 7 for descriptions of numbers, asterisks, and error bars.