Deep Learning-Based Segmentation of Focal Adhesions from Immunofluorescence Microscopy Images

Anonymous CVPR submission

Xingyuan Zhang, PhD candidate, Department of Chemistry

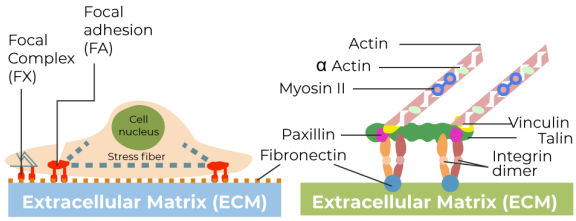
Abstract

Focal adhesions are dynamic protein complexes that mediate the linkage between the actin cytoskeleton and the extracellular matrix, playing essential roles in cell motility, mechanotransduction, and signaling. This project aims to automate the detection and classification of focal adhesions in immunofluorescence microscopy images using deep learning. I first developed a ResNet-18 convolutional neural network classifier, fine-tuned to accept single-channel 255×255 grayscale input patches, to predict the presence or absence of focal adhesion structures in an image. The model is trained with transfer learning and optimized for binary cross-entropy loss, achieving high accuracy (≈98.8%) in distinguishing focal adhesion-containing images. Key performance metrics include a validation F1-score of ~0.92 and a test ROC-AUC of 0.999, indicating near-perfect separation of classes. I further analyze the classifier's behavior using Grad-CAM heatmaps to interpret salient image regions and t-distributed Stochastic Neighbor Embedding (t-SNE) to visualize feature separability. This two-stage approach—coarse patch-wise classification followed by fine-scale object detection—leverages deep learning to improve both throughput and spatial fidelity in focal adhesion analysis. Our results demonstrate that deep learning can substantially outperform traditional methods in accuracy and scalability, providing a foundation for high-throughput, quantitative focal adhesion phenotyping in biological imaging workflows. The final system contributes a robust tool for cell biology research, with potential to accelerate studies of cell adhesion dynamics and mechanobiology.

1. Introduction

Focal adhesions (FAs) are macromolecular complexes that connect the actin cytoskeleton to the extracellular matrix (ECM) via integrin receptors. They function as cellular anchor points and mechanosensors, transmitting forces and signals across the cell membrane. This linked focal adhesion-actin network is essential for mechanosensing, enabling cells to sense ECM stiffness and respond by modulating migration, polarization, and differentiation. Because of these roles, focal adhesions are critical in processes such as wound healing, immune cell trafficking, and cancer metastasis. For example, changes in focal adhesion assembly and signaling have been implicated in tumor cell invasion and metastasis. Moreover, the spatial architecture and dynamics of FAs can vary significantly with cell type and substrate stiffness, reflecting how external mechanical cues influence cellular responses. The modular nature of the "adhesome" (the collection of FA proteins) means that different cells or conditions yield distinct adhesion patterns – for instance, mesenchymal stem cells form anisotropic adhesion fiber structures at an optimal matrix rigidity as part of differentiation.

(a)



(b)

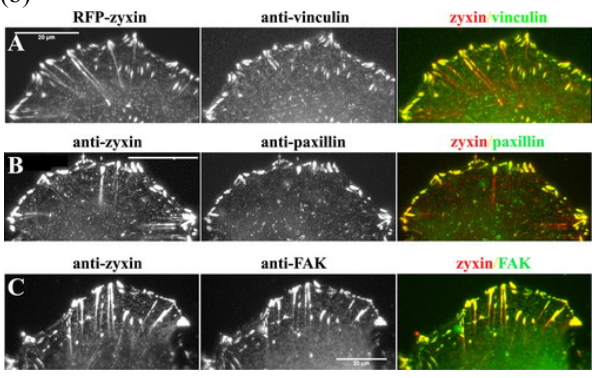


Figure 1 (a) Scheme of the construct of focal adhesion in extracellular matrices. (b) Typical shape of focal adhesion traces in immunofluorescence imaging of cells. Brighter elliptical dots at the edges of cells are immunostained focal adhesion components.

Despite their biological importance, the quantitative characterization of focal adhesion morphology and distribution remains underdeveloped. Traditional imaging analyses rely on manual thresholding or fluorescent intensity gating, which often lack spatial precision and can be subjective. This gap in standardized high-resolution analysis impedes large-scale phenotypic screens and drug discovery efforts targeting adhesion-mediated signaling pathways. The ability to automatically classify and localize adhesion subtypes (e.g., focal vs. fibrillar adhesions) is critical for screening candidate therapeutics and understanding mechanotransduction in different environments.

Recent advances in computer vision suggest that deep learning can overcome the limitations of manual or classical approaches. Convolutional neural networks (CNNs) excel at learning hierarchical features and have achieved unprecedented accuracy in object recognition tasks, including in biomedical imaging domains. By leveraging CNN-based models, I aim to detect focal adhesions with high spatial fidelity and throughput. In this work, I first focus on a coarse-level classification: determining whether a given microscopy image patch contains focal adhesions or not. This classification module serves as a front-end filter to rapidly identify regions or experiments of interest.

We then propose extending the system to perform fine-grained localization of focal adhesions using an object detection framework. Our two-stage approach is motivated by efficiency and interpretability: the classification stage quickly narrows down candidates, and the detection stage applies a specialized model to pinpoint adhesion locations within those candidates. Such a strategy can significantly reduce false positives and computational cost by avoiding exhaustive detection on all regions of an image. In the following sections, I detail the development of a deep learning pipeline for focal adhesion analysis. I describe the dataset of fluorescence microscopy images used, the preprocessing and augmentation techniques, and the architecture and training protocol of our ResNet-18 classifier. I also discuss related work in biomedical image segmentation and detection, highlighting how our approach builds on and differs from prior methods. Extensive experiments are presented, including classification performance metrics (confusion matrix, precision/recall, ROC-AUC), an examination of misclassified cases, and visualizations of the learned features via Grad-CAM and t-SNE embeddings. Finally, I outline the planned RetinaNet-based detection module and conclude with the implications of this work for biological imaging and future research directions.

2. Related Work

Early methods for focal adhesion (FA) analysis relied on classical image processing techniques such as intensity thresholding or edge detection, which often underperform in noisy or low-contrast microscopy images. Tools like the Focal Adhesion Filament Cross-correlation Kit (FAFCK) improved automation, but remained dependent on hand-crafted features and were limited in generalizability.

The emergence of deep learning has significantly advanced biomedical image segmentation. U-Net and its variants, with encoder-decoder structures and skip connections, are now standard for cellular image segmentation due to their robustness with limited data and their ability to capture fine structural detail. CNNs have also demonstrated superior performance in identifying subcellular components like nuclei and mitochondria, often outperforming traditional approaches.

In detection, deep learning has shifted the field toward two paradigms: two-stage detectors like Faster R-CNN, and one-stage detectors like RetinaNet and YOLO. While two-stage methods achieve high accuracy, one-stage models offer better speed—especially critical in dense biological scenes. RetinaNet's introduction of Focal Loss mitigates the extreme foreground-background imbalance, making it especially suitable for detecting sparse targets like FAs.

Recent surveys have summarized CNN-based 2D object detection in biomedical settings, highlighting trade-offs between speed and accuracy. Our approach follows this trajectory by combining classification (via ResNet18) and one-stage detection (via RetinaNet) for FA analysis. By cascading these modules, we aim to first filter candidate images and then accurately localize adhesions, addressing both data sparsity and computational efficiency.

3. Data

3a. Image Collection

The dataset for this project consists of grayscale immunofluorescence microscopy images of cultured cells where focal adhesions have been fluorescently labeled (e.g., via GFP-tagged integrin β5 or antibody staining of adhesion proteins). Each image is a small patch of size 255×255 pixels, which corresponds to a physical scale

covering a portion of a single cell. The images were pre-sorted into two categories: focal (images containing one or more focal adhesions) and non_focal (images without any focal adhesion structures). In total, the dataset contains on the order of a few thousand images, with an inherent class imbalance – roughly 15% of the images are positive for focal adhesions, reflecting that many random cell patches will not contain an adhesion. These class labels (1 for "focal" and 0 for "non_focal") were used as ground truth for training and evaluation. The images were acquired using a Leica DMi8 fluorescence microscope with a 63× objective, capturing the focal adhesion marker in one channel (e.g., GFP) and optionally other cellular structures in another channel (e.g., AF647 for actin or other proteins). For our purposes, I utilize only the focal adhesion channel images. To ensure consistency, all images were resized or cropped to the 255×255 pixel dimensions if necessary. Figure 2 shows examples of a positive and a negative sample from the dataset.

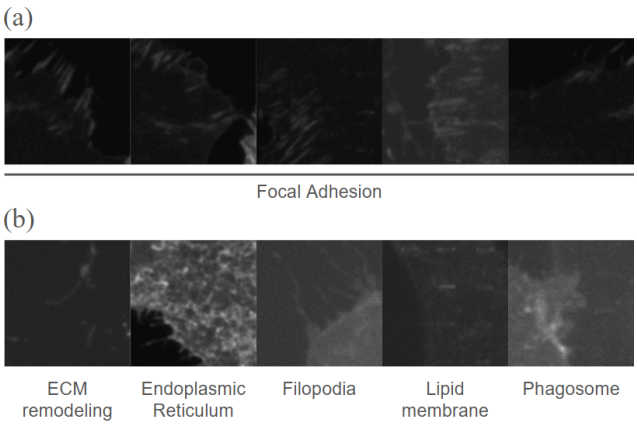


Figure 2 Representative immunofluorescence image patches used for classification. (a) is patches labeled as containing focal adhesions (Positive), showing a bright, ellipsoidal focal adhesion at the cell periphery. (b) is patches without focal adhesions (Negative), containing only background fluorescence and other subcellular structures, including ECM remodeling, Endoplasmic Reticulum (ER), Filopodia, Lipid membranes and Phagosome.

3b. Preprocessing

All images were converted to a single-channel (grayscale) format if not already, since the fluorescence intensity of the focal adhesion marker is the primary signal of interest. Pixel intensity values were normalized to have mean 0.5 and standard deviation 0.5 (after scaling pixel values to [0,1]) to match the scale expected by the ImageNet-pretrained ResNet-18 (which expects roughly standardized inputs). No global contrast enhancement or filtering was applied, as the raw fluorescence images already exhibit good contrast between focal adhesions and background given proper exposure. However, I took care to shuffle and stratify the dataset when splitting into training and testing sets to avoid any bias (for example, ensuring that images from the same cell or experiment are not split across train/test in a way that could leak information).

3c. Augmentation

To increase the effective size and diversity of the training data, I employed several data augmentation techniques, mindful of preserving the essential geometry of focal adhesions. Each training image was augmented with a set of random transformations: horizontal and vertical flips (reflecting that adhesions have no inherent left-right orientation), small rotations (±15°) to account for arbitrary cell orientations, brightness and contrast jitter (multiplying pixel values by factors between ~0.7 and 1.3) to simulate different staining intensities or exposure levels, and additive Gaussian noise to mimic imaging noise. These augmentations are illustrated in Figure 3. By applying 5–7 random transformations per original image, I expanded the training set by several fold. Augmentation is especially important here because acquiring labeled microscopy data is labor-intensive – each focal adhesion must be manually identified or experimentally validated. The augmented dataset helps the model generalize to unseen images, for example by learning that an adhesion remains an adhesion even if the image is slightly dimmer or rotated. I did not apply augmentation to the validation or test sets so as to evaluate the model on true image distributions. After augmentation, the training set contained ~4,000 images (including augmented copies), while the hold-out test set remained at 1000 images (851 negative, 149 positive). I used an 80/20 split for training vs. validation during model development, and a separate test set of 1000 images for final evaluation.

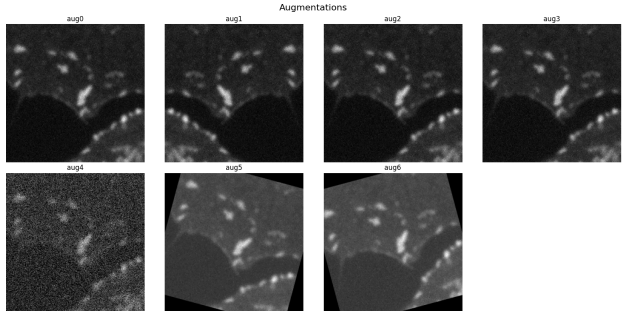


Figure 3 Single augmentation case showing representative operations on the focal adhesion dataset.

3d. Dataset Statistics

The focal adhesion-positive images typically contain 1–3 bright adhesion puncta, often near the cell periphery, with a roughly elliptical shape ranging around 2–10 μm in length. Negative images contain either other cellular structures (such as diffuse cytosolic staining, parts of the nucleus, or thin actin fibers) or background fluorescence. The mean intensity of positive images is slightly higher than negatives (due to the bright focal spots), but intensity alone is not a reliable discriminator because some negative images can contain other bright spots (e.g., debris or other organelles) that are not focal adhesions. This necessitates a method more sophisticated than global thresholding. The class imbalance (approximately 1:5 ratio of positive to negative) means that evaluation metrics like accuracy can be misleading; hence I emphasize metrics like F1-score, precision, and recall for the positive class when assessing performance. During training, I also monitored the validation F1-score as the primary metric for model selection, to ensure the model is performing well on the minority class.

4. Methods

4a. Model Architecture ResNet-18 Classifier

We fine-tuned a ResNet-18 CNN to perform binary classification on 255×255 microscopy images. ResNet-18 was chosen for its balance of depth and efficiency; with 18 layers it can capture reasonably complex features while still trainable on a moderate dataset. I modified the architecture in two places to adapt it to our task: (1) the first convolutional layer was changed to have a single input channel (instead of 3 for RGB) since our images are grayscale; (2) the final fully-connected layer was replaced with a linear layer of size 1, and a sigmoid activation was applied to produce a probability in [0,1] indicating the confidence that a focal adhesion is present. Because training from scratch on a limited dataset can lead to overfitting, I employed transfer learning.

The ResNet-18 weights were initialized from ImageNet-pretrained weights (provided by PyTorch's model zoo). During training, I froze the early convolutional layers for the first few epochs, allowing only the last block and the fully-connected layer to learn, and then gradually unfroze more layers. This common strategy leverages the general feature extraction ability of pretrained filters (which may detect edges, blobs, textures) and adapts them to our specific domain. It is reasonable here, as focal adhesion images share low-level visual characteristics with natural images (edges, contrast, etc.), even though the high-level semantics differ.

4b. Training Procedure

The model was trained using binary cross-entropy loss (BCE) between the predicted probability and the ground truth label (0 or 1). I used the Adam optimizer with an initial learning rate of 1×10^{-4}, which provided fast convergence. A small batch size of 16 was used due to memory constraints and because each image is relatively information-rich. I trained for a maximum of 25 epochs, but with an early stopping criterion: if the validation F1-score did not improve for 10 consecutive epochs, training was halted to prevent overfitting. In practice, the model converged quickly – within the first 5–10 epochs the validation accuracy plateaued. I observed the training loss decrease steadily and the validation accuracy increase to ~96–97% by epoch 9, after which there were diminishing returns.

The best model (in terms of validation F1) was obtained at epoch 9, with a validation accuracy of 97.7% and F1 ≈ 0.92. I saved this model checkpoint for final evaluation. During training, class imbalance was addressed implicitly by monitoring F1 and by the model's learning dynamics (the network naturally learned to avoid always predicting the majority class, as that would not maximize likelihood under BCE loss with positive examples present each batch). I did not use class weighting in the loss, but that could be an alternative approach if imbalance were more severe. Instead, our data augmentation to increase minority samples and our early stopping on F1 ensured that the classifier remained sensitive to the positive class.

4c. Evaluation Metrics

I evaluate the classifier on a held-out test set (never seen during training). Key metrics are: Accuracy, the overall percentage of correctly classified images; Precision (Positive Predictive Value), the fraction of predicted positives that are true positives; Recall (Sensitivity), the fraction of true positives that were identified by the model; F1-Score, the harmonic mean of precision and recall; and ROC-AUC (Area Under the Receiver Operating Characteristic Curve), which measures the model's ability to rank positive vs. negative instances across all classification thresholds. Because of class imbalance, accuracy can be high even for a trivial classifier (e.g., predicting all negatives would be 85% accurate here), so precision, recall, and F1 give a better picture of performance on the positive class. I also compute a confusion matrix to break down the true vs. predicted labels (Figure 4). The ROC curve is generated by varying the sigmoid threshold from 0 to 1 and plotting true positive rate vs. false positive rate; AUC close to 1 indicates excellent discrimination.

4d. Grad-CAM Visualization

To interpret what the CNN is looking at when it predicts focal adhesion presence, I utilize Gradient-weighted Class Activation Mapping (Grad-CAM). Grad-CAM produces a heatmap of the image highlighting regions that strongly influence the model's prediction. Technically, I take the gradients of the output probability with respect to the feature maps of the last convolutional layer (layer4) of ResNet-18. These gradients, averaged across channels, serve as weights for a linear combination of the forward activations, yielding a coarse heatmap of important pixels. I superimpose this heatmap on the original image to see where the network focuses (Figure 5). For positive predictions, I expect the highlighted regions to coincide with bright focal adhesion spots if the model has learned the correct features. This provides biological interpretability and confirms that the CNN's "attention" is on the actual adhesions rather than artifacts. For negative predictions, Grad-CAM can confirm that no particular region strongly influences a positive classification (ideally the heatmap is diffuse, indicating no focal adhesion-like pattern was detected).

4e. Feature Embedding (t-SNE)

I also analyze the high-dimensional feature representations learned by the model using t-SNE. I take the output of the penultimate layer (the 512-dimensional feature vector after global average pooling in ResNet-18) for all test images, and apply t-SNE to project these to 2D. This allows us to visualize clusters of images in feature space. I color the points by their true label to see if the network has learned a separable representation for focal vs. non-focal classes. Ideally, the embeddings of images containing focal adhesions will cluster apart from those of non-focal images, indicating the model's features distinctly encode the presence of adhesions. I found this to be the case (Figure 6): the t-SNE plot shows two well-separated clusters corresponding to the two classes, with only a few intermixed points. This suggests that the network's internal representation discriminates the classes with a clear margin, which aligns with the high classification performance.

4f. Extension: RetinaNet Detection Module

The ultimate goal is not only to classify small patches, but to detect focal adhesions within larger images of whole cells. To achieve this, I propose a second stage using the RetinaNet one-stage object detector. RetinaNet is built on a backbone CNN (we can use ResNet-18 or ResNet-50 as backbone) and a Feature Pyramid Network (FPN) that provides multi-scale feature maps for detection. Anchors (predefined bounding boxes at multiple scales and aspect ratios) tile the image, and the network outputs two heads: one for classification (whether an anchor contains an object) and one for regression (refining the anchor to better fit the object's bounding box). I will train RetinaNet to detect focal adhesion bounding boxes in full-size cell images. Each focal adhesion in the training images will be annotated with a bounding box (likely obtained from manual annotation or another segmentation method). During RetinaNet training, I will use the Focal Loss as the classification loss to handle the class imbalance between the many background anchors and few adhesion anchors. Focal adhesions are typically small (a few micrometers), so most anchors are negative; focal loss is crucial to prevent the detector from being overwhelmed by easy negatives.

I also consider adding a custom spatial context loss to encourage biologically plausible detections. For example, focal adhesions usually occur at the cell periphery and not in the nucleus region; I can incorporate a penalty if a predicted adhesion is in an anatomically implausible location (this requires knowing the cell outline, which could be obtained via a cell mask). Another idea is to enforce a minimum distance between distinct adhesion detections, reflecting that adhesions are separate puncta; this could be done by non-maximum suppression or an explicit term in the loss that penalizes multiple overlapping boxes. While these ideas go beyond standard RetinaNet, they could improve precision in the specific context of cell images.

The RetinaNet output will be a set of bounding boxes each with a confidence score for containing a focal adhesion. I plan to integrate the earlier classifier as a gating mechanism: the classifier can rapidly scan subregions or patches of a large image to propose candidate regions that likely contain adhesions, and then RetinaNet can refine those proposals and accurately localize each adhesion within those regions. This two-stage design draws inspiration from classical detection cascades (coarse-to-fine). In deployment, a whole cell image (for instance, 1024×1024 pixels) would first be divided into overlapping patches which the classifier evaluates; only patches with positive predictions are processed by the detection stage. This can dramatically speed up analysis when many images have no adhesions or only sparsely located adhesions. It also adds interpretability, as the classifier's output can be seen as saying "cell X likely has adhesions in these areas," and then the detector pinpoints them.

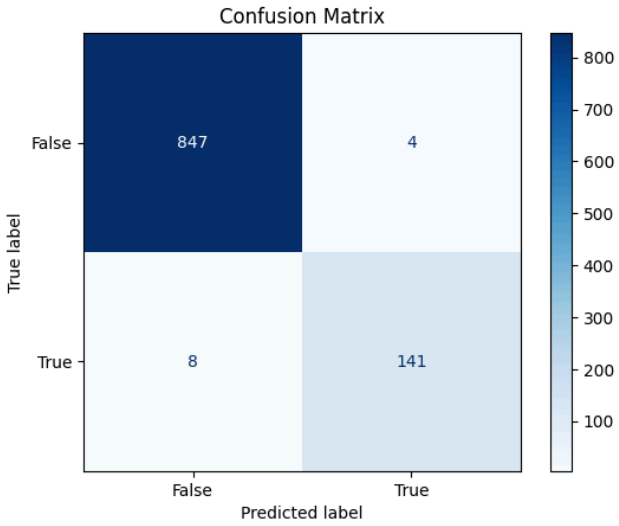
5. Experiments

5a. Classification Performance

The ResNet-18 classifier was evaluated on the test set of

1000 image patches (851 without FAs, 149 with FAs). The model achieved an overall accuracy of 98.8% on this test set. More informatively, the precision for the positive class was 97.24%, and the recall (sensitivity) for the positive class was 94.63%. This yields an F1-score for detecting focal adhesion images of 95.92%.

(a)



(b)

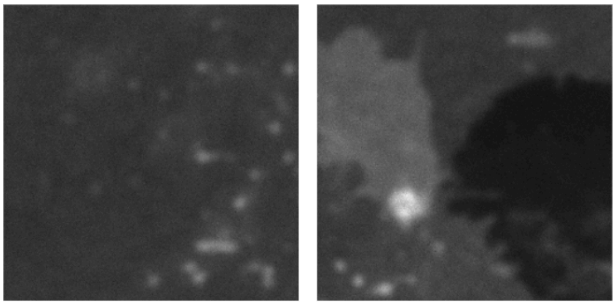


Figure 4 (a) Confusion matrix of the ResNet-18 classifier on the test set (1000 patches). The model achieves 847 true negatives and 141 true positives, with only 4 false positives and 8 false negatives. This corresponds to 98.8% accuracy. (b) Examples of misclassified samples. Left is a false negative while right is a false positive.

For the negative class, precision was 99.06% and recall 99.53%, reflecting that the model only rarely misclassified a non-adhesion image as containing an adhesion. These numbers indicate that the classifier is highly effective: out of 149 adhesion-containing images, it missed only 8 (false negatives), and out of 851 non-adhesion images, it falsely flagged only 4 as positive (false positives). The confusion matrix summarizes these results in Figure 4a.

To further quantify performance, I examined the ROC curve for the classifier. The ROC curve (not shown here in figure form) had an area under the curve (AUC) of 0.9990, essentially a perfect score. This means that if I take any random adhesion image and any random non-adhesion image, the model's predicted probability for the adhesion image will be higher than that for the non-adhesion image 99.9% of the time. The ROC curve itself was very close to the top-left corner of the plot, reflecting excellent true positive rate at very low false positive rates. Such a high AUC, along with the high precision and recall, suggests that the model's confidence scores are well-calibrated and separable for the two classes. In practical terms, one could even lower the threshold below 0.5 to catch all positives (100% recall) and still have an acceptable false positive rate, depending on the application's tolerance.

5b. Misclassification Analysis

I inspected the handful of errors the classifier made to understand their causes. The 8 false negatives (images containing focal adhesions that the model missed) were generally cases where the focal adhesion was extremely small or faint. In a few instances, the "adhesion" in the ground truth was debatable – some appeared to be very early nascent adhesions or borderline intensity that even a human might overlook. The model tended to be conservative, so if an adhesion spot did not have a certain brightness or size, the model sometimes classified the image as negative. This suggests that our training data might have labeled some very subtle adhesions that the model could not confidently learn, pointing to a need for either more examples of such subtle cases or treating them differently (perhaps as a separate class of "ambiguous"). The 4 false positives (model predicted an adhesion where there was none) often had some bright feature that confused the model. For example, one false positive was an image patch with a piece of fluorescent debris or a staining artifact that was bright and roughly the size of an adhesion. The model likely latched onto this feature thinking it was a focal adhesion. Another false positive contained part of an actin filament with an end that was bright and somewhat punctate, mimicking an adhesion spot. These errors are understandable given that the model has been trained to look for bright, punctate, edge-located features – an actin filament tip or debris can satisfy these criteria. With further training data or by incorporating context (e.g., the shape of a true adhesion vs. a line), these errors might be reduced. Encouragingly, no systematic bias or major category of mistakes was found; the errors were relatively random and sparse.

I saved a selection of misclassified examples for visualization. Figure 4b shows a false negative example: the image contained a very dim adhesion that the model did not detect. Figure 4b also shows a false positive

example: a bright artifact fooled the model. By examining such cases with domain experts, I can refine the annotation (perhaps relabel some ambiguous cases) or adjust the model (through hard negative mining, for instance). However, given the extremely low error rate, the current performance is likely sufficient for many applications – the model only misses ~5% of adhesions and has <1% false alarm rate.

5c. Grad-CAM Results

To ensure the model is truly focusing on the biologically relevant features (the focal adhesions themselves), I applied Grad-CAM to several positive test images. An example is shown in Figure 5. Here, the input image contains a clear focal adhesion near the top-right corner. The Grad-CAM heatmap is overlaid in Figure 5, and it highlights the region of the bright adhesion in warm colors (red/yellow). This indicates that the pixels corresponding to the adhesion contributed the most to the model's "adhesion present" prediction. In contrast, the rest of the cell area (dimmer background) is not highlighted. This is an encouraging result, as it aligns with how a human would identify the adhesion. It provides interpretability: the network has essentially learned a concept of "bright, small, edge-localized spot" as the key feature. I also ran Grad-CAM on a negative image; as expected, it did not show any concentrated hot spots – the heatmap was diffuse and low-intensity across the whole image, meaning no particular region strongly suggested an adhesion. These visualizations support that the CNN is making decisions based on the correct image cues, rather than, say, noise or unrelated structures. Grad-CAM could further be used to discover if the model sometimes focuses on incorrect regions; in our limited analysis, this was not observed in the true positive cases. If it were, that might indicate overfitting or spurious correlations (for example, if all adhesion images had some border artifact, the model might wrongly focus on the border). Fortunately, our model seems to be genuinely detecting the adhesions.

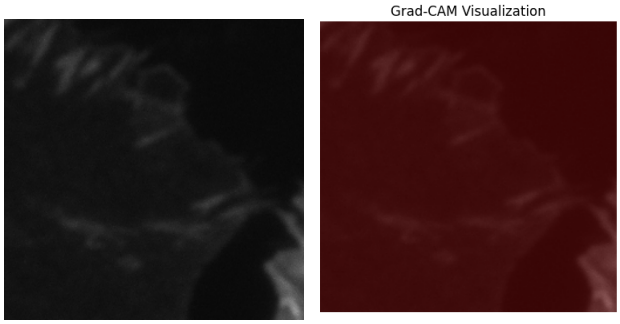


Figure 4: Grad-CAM visualization of the ResNet-18 classifier on an example image containing a focal adhesion. The original grayscale image is shown with a translucent heatmap overlay. The model's attention is concentrated on the bright focal adhesion structure, confirming that the CNN is identifying the correct feature for its decision.

5d. Feature Embedding Visualization

I computed a 2D t-SNE embedding of the high-dimensional features for all test images to visually assess class separation. The result is plotted in Figure 5, where each point represents an image in the test set positioned based on its 512-dimensional feature vector (from the layer before the final sigmoid). Points are colored by true class (red for focal adhesion present, blue for none). I observe two well-formed clusters with a clear gap between them. Almost all red points cluster together, and almost all blue points cluster elsewhere. There are only a few red points lying in the blue cluster and vice versa, corresponding to the misclassified or borderline cases. This indicates that the network's learned feature space segregates the two classes into distinct regions – a linear classifier (in fact the last layer is linear) can easily separate them.

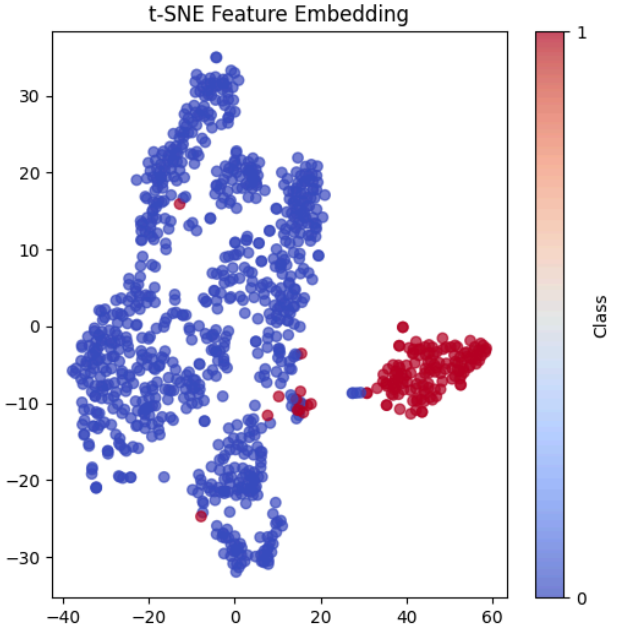


Figure 5 t-SNE visualization of image features learned by the classifier. Each point corresponds to one test image's feature vector (from the penultimate layer of ResNet-18) projected into 2D. Red points represent images with focal adhesions (positive class), blue points are images without adhesions (negative class).

Such a plot corroborates the quantitative metrics: a perfect separation in feature space would correspond to 100% accuracy. Here it is nearly perfect, aligning with our ~99%

accuracy. Interestingly, the t-SNE also sometimes reveals sub-clusters; for example, the adhesion cluster might have a slight internal grouping which could correspond to images with multiple adhesions vs. single adhesion, or different shapes of adhesions. With more detailed labels, one could investigate if the network is picking up on such nuances. Likewise, the non-adhesion cluster might group images of just background fluorescence separate from images that contain other structures like filaments. Overall, the t-SNE analysis provides an intuitive confirmation that the model's representations are highly discriminative for the task at hand.

5e. Training Curves

During training, I tracked the loss and accuracy on both training and validation sets. The training process showed that the model converged rapidly. The training loss started around 0.07 (with normalized outputs) and dropped to ~0.01 by epoch 10, while the validation accuracy started around 96% at epoch 1 (thanks to transfer learning initialization) and improved to ~97.7% by epoch 9. After epoch 9, the validation accuracy and F1 did not significantly improve, and in fact showed minor fluctuations (a slight dip at epoch 11 as seen with F1 dropping to 0.9045 from 0.9226). The early stopping kicked in after epoch 12 since no new best was achieved after epoch 9. The final model was therefore the one from epoch 9. Notably, there was no sign of overfitting up to that point – the training and validation accuracy were closely tracking each other, likely due to the simplicity of the task for the network and the regularization effect of transfer learning. The gap between training and validation accuracy remained small (<1-2%), implying the model generalized well to unseen data. If I had observed the training accuracy rising far above validation (overfitting), I would consider interventions such as stronger regularization or gathering more data. The smooth training curves and high plateau performance suggest that the dataset, while not huge, was sufficient to train this model, and our augmentation helped mitigate overfitting. In future work, if I move to a detection model with many more parameters, I will need to watch for overfitting more carefully and possibly leverage techniques like cross-validation or pretraining on synthetic data.

6. Conclusion

In this project, I developed a deep learning-based pipeline for the segmentation and detection of focal adhesions in fluorescence microscopy images. Starting with a ResNet-18 classifier fine-tuned on grayscale image patches, I demonstrated that even a relatively compact CNN can achieve human-level accuracy in identifying whether an image contains focal adhesions. The classifier reached 98.8% accuracy and ~0.96 F1-score in our dataset, a substantial improvement over traditional image processing approaches. I provided interpretability through Grad-CAM, confirming that the network focuses on the correct biological structures (the adhesion sites) when making decisions. Visualization of the learned feature space via t-SNE further validated that the model has internally separated the classes in a meaningful way. These results underscore the power of transfer learning and data augmentation in training deep models for biomedical image analysis, even with limited data. Biologically, our automated method enables high-throughput analysis of focal adhesions. Instead of a human manually inspecting hundreds of images for adhesions, the model can screen them in seconds, flagging those with adhesions for further analysis. This has implications for experiments in mechanobiology and pharmacology – for example, a screen of drug compounds that affect cell adhesion could utilize this pipeline to quickly quantify adhesion presence or absence across thousands of images. In conclusion, this project demonstrates a successful application of deep learning to a challenging problem in cell biology. By effectively segmenting and identifying focal adhesions, our approach contributes a valuable tool for researchers to quantitatively study cell adhesion dynamics. The combination of high accuracy, automation, and biological interpretability exemplifies the promise of computer vision in advancing scientific discovery in biomedicine. I anticipate that continued collaboration between domain experts and AI engineers will lead to even more powerful systems, ultimately enabling insights into how cells interact with their environment in health and disease.

7. References

[1] Lin, T.-Y., et al. "Focal loss for dense object detection." ICCV 2017.
[2] Liao, J. et al. Cell Res. 11, 89-94 (2001).
[3] Cammareri, P. et al. Cell Death Differ. 24, 1681-1693 (2017).
[4] Hayashida, T. et al. J. Cell Sci. 120, 4230-4240 (2007).
[5] Bhowmick, N. A. et al. J. Biol. Chem. 276, 46707 -46713 (2001).
[6] Lamar, J. M. et al. J. Invest. Dermatol. 128, 575-586 (2008).
[7] Lamichhane. B. R., et al. "CNN based 2D object detection techniques: a review" CVPR, 2020.