EVALUATING THE IMPACT OF IMAGE PREPROCESSING ON FOOD CLASSIFICATION: A COMPARISON OF LOW-LIGHT ENHANCEMENT, DEBLURRING, AND DOWNSCALING

Vincent Kao, Cheng-Yan Juang, Tien-Hao Chen

Department of Electrical and Computer Engineering
University of California, San Diego

ABSTRACT

This study discusses the impact of image preprocessing on a food type classification model. We use a standard computer vision model as our base, and apply a three kind preprocessing pipeline, including low-light enhancement, image deblurring, and downscaling. Each kind is evaluated with three representative algorithms. Using our own dataset, we compare classification accuracy in three scenarios: with unprocessed images on a pretrained model, with preprocessed images on the same pretrained model, and with unprocessed images on a model that is fine-tuned with new data we collected. Our results reveal which preprocessing strategies most benefit pretrained classifiers and whether spending effort on preprocessing or fine-tuning is more effective when adapting a model to new data.

***Index Terms*—** Image processing, low-light enhancement, deblurring, downscaling

1. INTRODUCTION

Food image classification plays a vital role in applications such as dietary evaluation to understand intake patterns and estimate nutrient levels. However, it faces several practical challenges. Real-world images are often affected by issues such as poor lighting, motion blur, and varying resolutions, which degrade visual quality and obscure discriminative features essential for accurate classification. These degradations introduce inconsistencies in brightness, sharpness, and scale, which reduce the robustness of deep learning models that rely on stable visual representations.

To address these challenges, our project investigates the impact of image preprocessing techniques across three domains: low-light enhancement, image deblurring, and downscaling for model compatibility. These algorithms are applied as preprocessing stages to enhance input quality and consistency prior to classification. We evaluate their effects on the performance of a SigLIP model fine-tuned on the Food-101 dataset, aiming to determine how different enhancement strategies contribute to more robust and accurate food recognition.

2. RELATED WORKS

We reviewed related works in three significant image processing directions, which are low-light enhancement, image deblurring, and downscaling. Each direction represents an aspect of image quality degradation that can significantly affect the performance of the computer vision model. Through the analysis of representative research papers that applied algorithms within these domains, we summarized the key methodologies, strengths, and limitations of the algorithms used in each category.

2.1. Low-light Enhancement

Low-light conditions often degrade image quality and hinder the performance of classification models. Poor illumination reduces contrast and obscures important features, which makes it harder for models to extract meaningful patterns. To address this issue, various image enhancement techniques aim to improve visibility and contrast before classification. Some commonly used approaches are gamma transformation, adaptive histogram equalization (AHE), contrast-limited adaptive histogram equalization (CLAHE), and Retinex-based enhancement. Each of them offer different trade-offs between simplicity, adaptiveness, and perceptual quality.

2.1.1. Gamma transformation

Gamma transformation is a nonlinear intensity mapping that adjusts brightness according to a power-law function. By setting the gamma value less than 1, dark regions can be brightened while maintaining a smooth tone distribution. The method is simple, fast, and suitable for global brightness correction, but it does not handle uneven lighting well. Adaptive variants improve this by calculating gamma based on image statistics. For example, Huang et al. (2013)[1] proposed an adaptive gamma correction using the cumulative intensity distribution to brighten underexposed images without overexposing bright regions.

2.1.2. Adaptive Histogram Equalization (AHE)

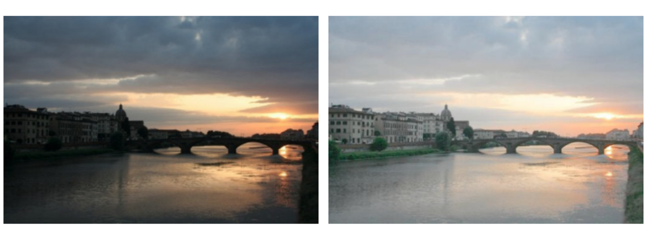
AHE enhances image contrast by computing several histograms corresponding to distinct regions of the image and using them to redistribute the intensity values of the image locally. Unlike global histogram equalization, which applies a single transformation to the entire image, AHE adapts to local intensity variations, making it highly effective for revealing details in both dark and bright regions simultaneously. This adaptability improves visibility in images with non-uniform illumination or low contrast. However, AHE tends to over-amplify noise in relatively homogeneous areas, as local contrast enhancement can exaggerate minor intensity fluctuations. It also involves a higher computational complexity compared to global methods, due to the need for processing per region.

2.1.3. Contrast-Limited Adaptive Histogram Equalization (CLAHE)

CLAHE enhances local contrast by dividing an image into small tiles and equalizing the histogram of each region, while limiting noise amplification. This approach reveals hidden details in shadowed areas and improves feature visibility for recognition systems. Compared with global methods, CLAHE produces better local contrast and edge clarity but requires careful parameter selection to avoid artifacts. A review of CLAHE-based methods in face recognition has shown its effectiveness in improving visual quality and recognition performance under poor lighting conditions[2].

2.1.4. Single-Scale Retinex (SSR)

Retinex-based methods, inspired by human vision, model an image as the product of illumination and reflectance. The Single-Scale Retinex (SSR) algorithm estimates illumination using a Gaussian filter and enhances the reflectance to recover detail from dark areas while preserving overall color balance. SSR handles non-uniform lighting more effectively than global methods, although it requires more intensive computation and is sensitive to parameter settings. Liu et al. (2021) presented a fast Retinex-based approach that improves visibility and color consistency in low-light images[3].



Before (left) and after (right) applying SSR

In summary, these three enhancement techniques represent strategies for improving image quality before classification. Together, they offer a comprehensive foundation for evaluating how classical image processing can improve classification performance under low-light conditions.

2.2. Image Deblurring

Image deblurring has been studied for decades and evolves from classical signal processing models to modern deep learning frameworks. In this report, we refer to the paper by Li [4], which categorizes existing image deblurring research into two main directions, which are traditional model-based methods and deep learning–based methods. To better understand the strengths and weaknesses of each category, a representative algorithm and the work chosen from each category are briefly discussed below.

2.2.1. Traditional Model-Based Method - Wiener Filter

Early studies focused on modeling the blurring process mathematically and restoring images through optimization or frequency-domain filtering. They are primarily designed for non-blind deblurring, assuming that the blur kernel is known. However, since they do not fully utilize the prior statistical information of natural images, their restorations often have inaccurate, oversmoothed edges and lose texture detail. A classic example is the Wiener Filter, which reconstructs images by minimizing the mean square error in the frequency domain.

It assumes that the degradation process can be represented as a convolution of the ideal image with a blur kernel, corrupted by additive noise. Mathematically, the Wiener filter reconstructs the image in the Fourier domain according to:

$$\hat{F}(u,v) = \frac{H^*(u,v)}{|H(u,v)|^2 + \frac{S_n(u,v)}{S_f(u,v)}} G(u,v) \tag{1}$$

where $H(u,v)$ is the point spread function (PSF), $S_n(u,v)$ and $S_f(u,v)$ are the noise and image power spectra, and $G(u,v)$ is the degraded image.

By weighting the inverse of the PSF according to the noise-to-signal ratio, the Wiener filter effectively balances deblurring and noise suppression[5]. According to Bojarczak and Lukasik[5], the Wiener filter yields excellent results when the variance and type of noise are known.

In their experiments, when the filter's noise variance was adjusted correctly, the reconstructed image achieved a high improvement in Signal-to-Noise Ratio (ISNR = 1.55 dB). However, when the assumed noise variance was mismatched, the performance dropped drastically (ISNR = –0.13 dB), resulting in visible artifacts and excessive smoothing. This shows that Wiener filtering is highly sensitive to the accuracy of its parameters.

The main limitation highlighted in the paper is that real-world blurring cases rarely provides knowledge of the noise statistics or blur kernel. In real-world cases, when the blur kernel varies across the image or when the noise spectrum is unknown, Wiener filtering may amplify high-frequency noise or generate ringing artifacts. Nevertheless, its simplicity, strong theoretical foundation, and computational efficiency make it a valuable benchmark for evaluating more advanced restoration methods.

In conclusion, Wiener Filter performs best when the blur kernel and noise parameters are known. However, it is susceptible to noise and produces ringing artifacts when these parameters are inaccurate.

2.2.2. *Deep learning based methods - MSCNN*

With the rise of convolutional neural networks, image deblurring has transitioned to a data-driven approach. A representative model is the Multi-Scale Convolutional Neural Network (MSCNN), which restores images from rough to fine resolutions. Seungjun Nah, Tae Hyun Kim and Kyoung Mu Lee proposed the Deep Multi-Scale Convolutional Neural Network (MSCNN) [6], a pioneering end-to-end framework for blind deblurring in dynamic scenes.

Unlike traditional optimization-based methods that rely on explicit blur kernel estimation, MSCNN directly restores a sharp image from a single blurry input without any prior knowledge of motion or camera parameters. The network adopts a multi-scale coarse-to-fine architecture, which mimics the structure of classical deblurring pipelines.

Each scale predicts a progressively sharper image, and a multi-scale loss function ensures that intermediate outputs at lower resolutions contribute to the final refinement. This hierarchical design allows the model to handle spatially varying and non-uniform motion blur effectively. MSCNN eliminates the need for kernel estimation, avoids ringing artifacts, and generalizes well to complex motion and dynamic scenes.

However, the model requires large paired datasets, extensive GPU training, and does not generalize well to unseen blur types without retraining. Furthermore, the multi-scale design increases computational cost, making it less practical for real-time applications.

In conclusion, it handles large motion blur effectively but requires extensive training data and high computational resources.

2.3. Downscaling

Downscaling is an important preprocessing step before feeding images into a computer vision classification model, as most modern neural architectures, including the one adopted in this work, require inputs of size 224×224. In contrast, raw images captured by smartphones and cameras are much larger. There are multiple approaches to downscale an image, and they can be broadly categorized as traditional interpolation, optimization-based methods, and deep learning-based methods.

2.3.1. *Traditional Interpolation*

Commonly used traditional interpolation methods include bilinear, bicubic, and Lanczos, each relying on a fixed convolution kernel. The kernel defines the weights assigned to pixels in a local neighborhood when computing each output value. The primary distinction between these methods lies in how the weights are determined. Bicubic interpolation computes each output pixel from a 4×4 pixel region to produce smooth results, while Lanczos uses a windowed sinc function over a larger domain, and yields sharper details and better suppression of aliasing in general. Importantly, once the kernel weights are set, they remain uniform throughout the image. The primary drawback of these approaches is their inability to adapt to diverse image content, which often results in the loss of structural detail, edge blurring, and the appearance of artifacts such as aliasing or ringing, particularly in areas with complex textures or sharp transitions.

2.3.2. *Deep Learning Methods*

To address these limitations, deep learning methods have been introduced, such as CNN-based models like the Scale-arbitrary Invertible Image Downscaling Network (AIDN) [7] and GAN-based models like KernelGAN [8]. These approaches can learn to synthesize content-aware, spatially varying kernels for each image patch or region and allows for superior retention of details and textures compared to fixed-kernel interpolation. However, their effectiveness relies on training with a large, diverse dataset of annotated images, and their adaptability is restricted without retraining when new data distributions are encountered. While deep learning-based downscaling excels in generalization and can be seamlessly integrated into modern vision pipelines, the high demands for data and computation are significant trade-offs.

2.3.3. *Optimization-Based Methods*

The advantage of Optimization over Deep Learning methods is that they do not need large-scale training data, while allowing content-adaptive refinement. Instead of globally fixed kernels or dataset-driven learning, these algorithms iteratively adjust the value of each pixel, typically starting from the output of a traditional interpolation method. Examples include Content-Adaptive Image Downscaling [9], Perceptually Based Downscaling of Images [10], and L0-regularized optimization [11]. The goal is to optimize pixel intensities or local filtering operations to maximize criteria such as edge sharpness, structural similarity (SSIM), or the sparsity of image gradients, thereby enhancing the perceptual quality of each input individually. The strength of optimization-based downscaling lies in its capacity for image-specific adaptation and the practical preservation of relevant details. However, this comes at the cost of increased computational load, as the underlying iterative procedures are resource-intensive, especially for high-resolution content.

3. PROPOSED WORK

The goal of this project is to investigate the impact of various image preprocessing techniques on the performance of a classifier in the presence of low-quality images. Our classifier is a SigLIP model fine-tuned on the Food-101 dataset, and we have a small dataset of low-quality images for evaluation. The images will first be preprocessed by our algorithms before being fed to the classification model.

3.1. Dataset

The dataset used in this study was designed to reflect the most common image degradation issues encountered when using a smartphone for food recognition. We collect the food image set consisting of various dishes—such as ramen, curry rice, salad, etc., captured under diverse conditions. These images simulate typical degradation scenarios found in mobile photography, including motion Blur, caused by hand-held camera shake or rapid shooting, which results in blurred edges and the loss of fine texture, and Low-Light and Noise, occurring in poorly lit restaurants where high ISO introduces random sensor noise and long exposure further exacerbates motion blur. These variations collectively create a challenging yet realistic dataset for evaluating image enhancement and deblurring algorithms. The purpose of this study is to analyze how different types of image degradation affect the performance of a computer vision model, and how classical enhancement algorithms can mitigate these effects.

3.2. Image Processing Algorithms

The following three subsections will include the algorithms we propose to use to preprocess our images in each category.

3.2.1. Low-light enhancement

To improve brightness and contrast in images for better classification performance, we will apply and compare three image processing algorithms: Gamma Transformation, CLAHE, and SSR. These algorithms were selected for their distinct strengths: Gamma Transformation provides efficient global intensity correction; CLAHE enhances local contrast while minimizing noise amplification; and SSR models illumination and reflectance to restore visual balance under uneven lighting. By integrating these enhancement methods into the preprocessing pipeline, our objective is to improve the ability of the classification model to extract discriminative features from dark or underexposed images, increasing precision under low-light conditions. The choice of classical, training-free algorithms ensures interpretability, reproducibility, and computational efficiency, making this enhancement stage a practical and explainable component of the overall classification system.

3.2.2. Image Deblurring

While existing deblurring research primarily targets motion-induced blur through complex kernel estimation or deep-learning-based restoration, these approaches are unsuitable for lightweight, real-world applications, such as mobile image classification. In contrast, this project focuses on improving overall image quality—including edge sharpness, contrast, and noise reduction—through simple, interpretable, and computationally efficient filtering techniques. Thus, three representative algorithms are implemented: the Simple Sharpening Kernel, Unsharp Masking (USM), and a Simplified Wiener Filter. The sharpening kernel operates in the spatial domain using a fixed high-pass convolution mask to emphasize edges and fine details, effectively improving perceived sharpness and local contrast.

The USM method enhances detail visibility by subtracting a blurred version of the image from the original and adding the weighted difference back, expressed as

$$I_{\text{USM}} = I_{\text{original}} + k(I_{\text{original}} - I_{\text{blurred}}) \quad (2)$$

Where k controls the enhancement strength, this approach allows greater flexibility than the fixed kernel and is particularly effective for improving texture-rich regions. The Wiener Filter, initially designed for frequency-domain deblurring, is implemented here in a simplified form as an adaptive denoising tool. Because the actual blur kernel and the signal-to-noise ratio are unknown in our real-world dataset, the complete deconvolution formula introduced in the related works[5] cannot be applied. Instead, the adaptive filtering capability of the simplified Wiener algorithm is used to locally estimate and reduce image variance, suppressing random noise while preserving overall structure.

The Simple Sharpening Kernel and Unsharp Masking (USM) are applied to restore edge definition in images that are blurred or have low contrast. At the same time, the Simplified Wiener Filter focuses on suppressing sensor noise in high-ISO or low-light conditions. For each image, three enhanced versions are generated and evaluated using the pretrained model trained on a clean dataset. Classification accuracy serves as the primary evaluation metric, directly reflecting the model's recognition ability rather than perceptual image quality.

Through this comparison, the analysis aims to determine which type of degradation results in a greater performance loss for our classification model.

3.2.3. *Downscaling*

We selected one algorithm from each major downscaling category, including traditional interpolation, optimization-based, and deep learning, for experimental comparison, as each provides unique trade-offs. For traditional interpolation, we use Lanczos resampling with anti-aliasing due to its efficiency and ability to retain structural details during significant reductions (e.g., 4000×3000 to 224×224). Deep learning methods, specifically CNN-based kernel resampling, are preferred over GAN-based approaches since GANs tend to generate unrealistic textures unsuitable for aggressive downscaling. In contrast, CNN-based models learn optimal content-aware kernels but require substantial training data. For optimization-based techniques, Content-Adaptive Image Downscaling (CAID) was chosen for its effective edge and artifact management by locally adapting interpolation kernels. In contrast, L0-regularization emphasizes strong edge preservation but can be computationally heavy and sometimes over-sharpen subtle textures.

3.3. Evaluation Setup

The proposed work consists of the following four main parts, in which we will compare the respective results.

3.3.1. *Part I: Baseline Evaluation*

To check the model's baseline performance, We will first use our bad-quality images as the direct input to the classification model.

3.3.2. *Part II: Individual Evaluation*

For each preprocessing category, the algorithms will be tested independently. Each algorithm will be applied individually to the images, and the resulting images will be fed into the classifier to measure classification accuracy. The algorithm of each category that achieves the highest accuracy will be selected as the best-performing method.

Once the best algorithm from each category is found, these selected methods will be combined in a single preprocessing pipeline. The classifier's performance on the low-quality images with the combined preprocessing will then be compared with the baseline performance without any preprocessing. This step will allow us to evaluate the benefit of sequentially applying the best algorithms across different preprocessing categories.

Example:

- Low-light enhancement: L1, L2, L3 $\rightarrow$ L2 performs best
- Deblurring: D1, D2, D3 $\rightarrow$ D1 performs best
- Downscaling: S1, S2, S3 $\rightarrow$ S2 performs best

The combined pipeline L2 + D1 + S2 will then be applied to the image before the base model.

3.3.3. *Part III: Combination Search*

After identifying the best-performing algorithms in Part II, we will perform an exhaustive search of all possible algorithm combinations across the three preprocessing categories. We will loop through every combination of algorithms to the images and measure the accuracy for each combination. For example, with three algorithms per category, this would result in $3 \times 3 \times 3 = 27$ combinations.

The purpose of this phase is to determine whether a combination of algorithms other than the best ones can provide superior classification performance. The best combination from this search will be compared to the baseline model and the Phase I results to identify the most effective preprocessing strategy.

3.3.4. *Part IV: Model Fine-Tuning on Low-Quality Images*

In the last part, the SigLIP model will be fine-tuned directly on the low-quality image dataset. The performance of the fine-tuned model will be compared with the preprocessing pipelines identified in Phases I and II, as well as the base model trained on standard-quality images. This part would allow us to determine the effectiveness of model learning versus image preprocessing to improve classification performance on low-quality images.

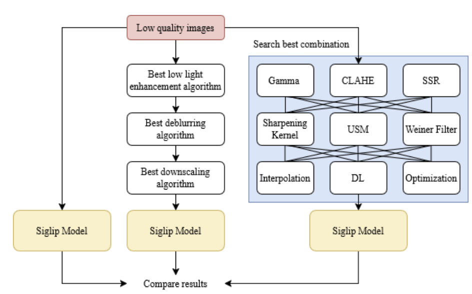


Fig. 1. Diagram of Part I to Part III, from left to right

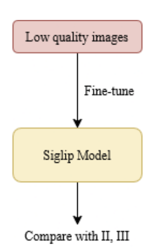


Fig. 2. Diagram of Part IV

3.4. Expected Outcome

Generally, the performance of the food classification model is expected to vary depending on the type of image degradation and the preprocessing algorithm used.

Among the low-light enhancement techniques, Retinex-based methods are expected to have the best results, as they separate illumination from reflectance, resulting in natural brightness correction and consistent color balance. The gamma correction and CLAHE on the other hand, may over-amplify highlights or increase noise in dark regions. For image deblurring, classical sharpening and Wiener filtering are expected to moderately improve recognition accuracy under mild blur but may introduce ringing artifacts or oversharpening effects when blur is severe. In downscaling-related preprocessing, optimization-based approaches such as CAID and CNN-based methods like AIDN are expected to outperform traditional interpolation algorithms (e.g., Lanczos) by restoring high-frequency textures and preserving fine food details critical for classification.

Overall, low-light enhancement and high-quality downscaling are expected to contribute more to classification performance than traditional deblurring. This is because Vision Transformer–based models, such as SigLIP, are more sensitive to global illumination and color balance than to moderate spatial blur, due to their attention-based global feature representation.

In the final part, the SigLIP model will be fine-tuned directly with the low-quality image dataset to examine whether model adaptation can surpass the effectiveness of preprocessing. Fine-tuning is expected to yield the highest accuracy by enabling the model to learn feature representations that are inherently robust to blur, noise, and illumination variations.

However, this improvement may come at the cost of higher computational cost and reduced generalization to unseen degradation types, as the model becomes specialized to the fine-tuned dataset. In contrast, classical enhancement techniques are model-agnostic and computationally lightweight, providing broader applicability despite their limited ability to recover fine details.

4. CONCLUSION

This project aims to evaluate the impact of algorithmic image preprocessing techniques, including low-light enhancement, deblurring, and downscaling, on the performance of food type classification.

By establishing a controlled experimental pipeline and testing algorithms within each category, we aim to identify the preprocessing strategies that most effectively enhance model performance under real-world degradations. The inclusion of both algorithmic comparison and model fine-tuning allows for an assessment of whether it is more beneficial to enhance the input data or adapt the model itself when dealing with low-quality images. The outcome gives us a better understanding of how various preprocessing techniques affect classification accuracy under degraded imaging conditions.

On balance, the study will provide comparative evidence on the relative benefits of preprocessing versus fine-tuning, supporting more informed decisions when deploying image classifiers in practical, low-quality environments.

5. REFERENCES

[1] Yi-Sheng Chiu, "Efficient contrast enhancement using adaptive gamma correction and cumulative intensity distribution," in *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*. IEEE, October 2011.

[2] P. Musa, "A review: Contrast-limited adaptive histogram equalization (clahe) methods to help the application of face recognition," in *Proceedings of the 2018 Third International Conference on Informatics and Computing (ICIC)*. IEEE, 2018, pp. 1–6.

[3] Mahdi Mirzapour, Ali Keshavarz Nasab, Amir Movafeghi, and Effat Yahaghi, "Retinex theory based automated contrast enhancement of gamma radiographic images of pipe welds," *Journal of Nondestructive Evaluation*, vol. 44, 06 2025.

[4] C. Li, "A survey on image deblurring," in *Proceedings of the International Conference on Image and Signal Processing*. IEEE, 2022, pp. 1–6.

[5] P. Bojarczak and Z. Lukasik, "Image deblurring – wiener filter versus tsvd approach," *Advances in Electrical and Electronic Engineering*, vol. 6, no. 2, pp. 85–91, 2007.

[6] S. Nah, T. H. Kim, and K. M. Lee, "Deep multi-scale convolutional neural network for dynamic scene deblurring," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, 2016, pp. 3883–3891.

[7] Jinbo Xing, Wenbo Hu, and Tien-Tsin Wong, "Scale-arbitrary invertible image downscaling," *arXiv preprint arXiv:2201.12576*, 2022.

[8] Sefi Bell-Kligler, Assaf Shocher, and Michal Irani, "Blind super-resolution kernel estimation using an internal-gan," *arXiv preprint arXiv:1909.06581*, 2019.

[9] Johannes Kopf, Ariel Shamir, and Pieter Peers, "Content-adaptive image downscaling," *ACM Transactions on Graphics*, vol. 32, no. 6, Nov. 2013.

[10] A. Cengiz Öztireli and Markus Gross, "Perceptually based downscaling of images," *ACM Transactions on Graphics*, vol. 34, no. 4, Aug. 2015.

[11] Junjie Liu, Shengfeng He, and Rynson W. H. Lau, "L0-regularized image downscaling," *IEEE Transactions on Image Processing*, vol. 27, no. 3, pp. 1076–1085, Mar. 2018.