

FDWST: Fingerphoto Deblurring using Wavelet Style Transfer

David Keaton, Amol S. Joshi, Jeremy Dawson, Nasser M. Nasrabadi

West Virginia University

{dck0013, asj00003}@mix.wvu.edu, {jeremy.dawson, nasser.nasrabadi}@mail.wvu.edu

Abstract

The challenge of deblurring fingerphoto images, or generating a sharp fingerphoto from a given blurry one, is a significant problem in the realm of computer vision. To address this problem, we propose a fingerphoto deblurring architecture referred to as Fingerphoto Deblurring using Wavelet Style Transfer (FDWST), which aims to utilize the information transmission of Style Transfer techniques to deblur fingerphotos. Additionally, we incorporate the Discrete Wavelet Transform (DWT) for its ability to split images into different frequency bands. By combining these two techniques, we can perform Style Transfer over a wide array of wavelet frequency bands, thereby increasing the quality and variety of sharpness information transferred from sharp to blurry images. Using this technique, our model was able to drastically increase the quality of the generated fingerphotos compared to their originals, and achieve a peak matching accuracy of 0.9907 when tasked with matching a deblurred fingerphoto to its sharp counterpart, outperforming multiple other state-of-the-art deblurring and style transfer techniques.

1. Introduction

The task of deblurring images has long been an open problem in the field of computer vision. When an image is randomly blurred, it can become very difficult to render the high-frequency components of the image. Consequently, the captured image predominantly contains lower frequencies. Fingerprint photographs, or fingerphotos, have become more popular in recent years. The ease with which a fingerphoto can be acquired is much higher than that of contact-based fingerprints. However, this ease of acquisition comes with a downside, as fingerphotos are much more subject to accidental blurring than contact-based prints. The effects of motion blur, poor camera quality, and even unstable hands can be detrimental to the quality of fingerphotos.

Conceptually, any given image can be construed as being made up of sets of information at different frequencies, ranging from low to high. Large, monochrome, or texture-

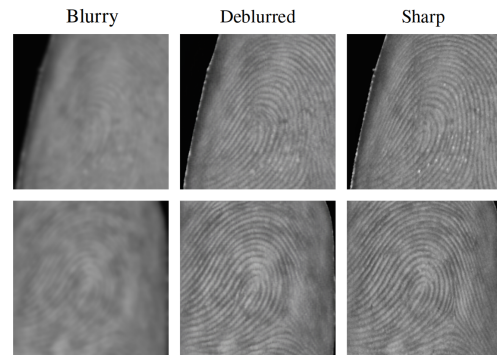


Figure 1. Sample pairs of sharp and blurry fingerphotos, with the deblurred version in the middle column.

like areas are primarily made up of low-frequency information. These features have minimal variety in their pixel values and, therefore, are less prone to blurring. At the other end of the spectrum, high-frequency information refers to the edges, lines, and points of an image; sections that display a rapid transition in pixel value over a small change in location. This information is vulnerable to being lost due to blurring as the majority of pixels surrounding and making up an edge or line have drastically different values from one another. This causes high-frequency information to be dispersed into its neighboring pixels and, therefore, destroyed.

The inherent uncertainty involved in deblurring, mainly in the restoration of lost high-frequency information, has proven to be a daunting task for researchers. The process of recovering lost high-frequency detail is generally considered an ill-posed problem. This arises from the fact that there is no deterministic way to guarantee that a deblurred image corresponds to the unique original of an arbitrarily blurred image. The nature of fingerprints, mainly their seemingly random arrangement of ridges, loops, and bifurcations, makes them particularly formidable for classical deblurring techniques to effectively enhance. Unfortunately, these features, referred to as minutiae, comprise the bulk of the important identification information in any given fingerprint. Modern fingerprint identification systems, such as Neurotechnology's Verifinger SDK or Innovatrics' ID-

Kit SDK [20, 11], heavily rely on the properties of these minutiae for their fingerprint-subject verification, in addition to several other proprietary algorithms. These systems look mainly at the direction, ridge orientation, thickness, and locations of minutiae points, all of which are highly susceptible to loss through blurring. During the deblurring process, it is imperative that these identity-related features are reconstructed in a way that accurately reflects the originals, as even minor alterations to minutiae can significantly impact identification accuracy.

In this paper, we propose a modified deep convolutional neural network termed Fingerphoto Deblurring using Wavelet Style Transfer (FDWST), which aims to combine the best aspects of both style transfer techniques and the Discrete Wavelet Transform (DWT). Our base model leverages a standard autoencoder, although with one significant modification. Unlike conventional autoencoders, we entirely discard the encoder and replace it with the DWT operation, allowing our input images to be split into distinct wavelet frequency bands. Additionally, following the DWT, we perform our Wavelet Style Transfer Module, facilitating style transfer within the wavelet domain instead of in the feature domain. By substituting the encoder for the DWT, we force our inputs to be split into distinct frequency bands, thereby enabling a more adequate transfer of sharpness information at different frequency bands. This strategy improves the overall quality of our deblurred images.

Our proposed model goes one step further in that it can deblur any given fingerphoto with a single style image. Specifically, for every blurry fingerphoto, a single sharp fingerphoto is used. This fingerphoto undergoes binarization, i.e., every value of the pixel is set to either 0 or 1, enhancing the quality of the sharpness information. Further details on this technique are elaborated in Section 3.3. Some example results from our model are displayed in Fig. 1. Our primary contributions are outlined as follows:

- Integration of Style Transfer and DWT in a novel approach with the goal of effectively transferring sharpness and clarity information to deblur fingerphotos.
- Exhibit how Wavelet Style Transfer enhances fingerphoto deblurring by facilitating the transfer of high-frequency sharpness information between images.
- Presentation of qualitative and quantitative evidence showcasing the superior performance of our model over other state-of-the-art deblurring and style transfer architectures.

This paper is organized into seven main sections. Section 2 discusses related work in the fields of deblurring and style transfer, as well as some other wavelet-based style transfer methods. Section 3 explains some background information needed for the model and some details on certain decisions

regarding our design choices. Lastly, Sections 4, 5, 6 and 7 give a thorough overview of our implemented architecture, experiments, evaluation results, and ablation study.

2. Related Works

2.1. Image Deblurring

Image deblurring has been a fundamental task in the field of computer vision and digital graphics for many years [19, 22, 23, 27, 8]. The objective is straightforward; given a blurry image I_b that has been corrupted by a blur kernel K and unknown noise N , restore a sharp image I_s devoid of both the effect of K , as well as N . There is irony in the fact that this simple concept does not have a deterministic solution, as there is inherent ambiguity involved when removing blur from an image. Despite this, models that both know (non-blind) and do not know (blind) the blur kernel have made impressive headway in recent years. Kupyn *et al.* [15] use a GAN architecture and simulate realistic motion blur to reverse the effects of motion blur in natural images. Chiu and Gurari [26] introduce an interesting approach, involving two GANs, one of which learns to blur and the other to deblur, and they compete with each other in such a way that fake deblurred images cannot be determined from real sharp ones. Since these deblurred images cannot be distinguished from real sharp ones, they are, by proxy, sharp themselves. Joshi *et al.* [13] incorporate multi-stage GAN architecture to deblur fingerphotos. However, the convoluted network requires a specific resolution to operate, making it challenging to use in a real-world scenario.

2.2. Neural Style Transfer

Style transfer refers to the task of amalgamating two images, the content image and the style image. The objective is to combine them in such a way that the overall structure of the content image is retained, while reconstructing it in the same artistic essence as the style image. Generally, style transfer architectures are designed with the goal of mimicking an art, and not for a more practical task such as deblurring. Beginning with the landmark work of Gatys *et al.* [7] in 2015, neural style transfer has seen an explosion of interest recently, [9, 18, 14, 21] with the growing popularity of artificially generated artwork. Gatys *et al.* introduced style transfer by utilizing the Gram matrix to perform a form of style loss between two images. Huang and Belongie [10] pioneered a new approach, utilizing Adaptive Instance Normalization, aligning the mean and standard deviation of the content image with the style image and thereby transferring the style. [17, 21] expanded upon the ideas of [10], introducing attention and a form of weighted mean-variance maps, respectively. Deng *et al.* [5] make use of transformers to generate domain-specific content and style sequences.

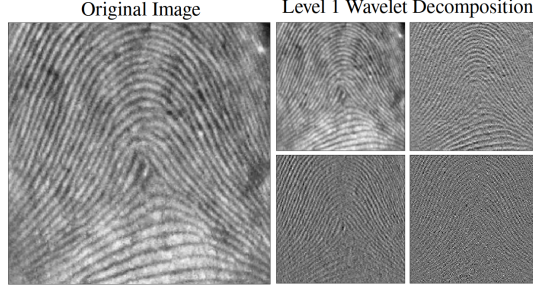


Figure 2. A sharp fingerphoto alongside its corresponding level 1 wavelet decomposition. The top left sub-band is an approximation of the original image, being a smaller smoothed version. The remaining three sub-bands contain horizontal, vertical, and diagonal information, respectively.

2.3. Wavelet-Infused Networks

The use of wavelets in style transfer has seen very little exploration in recent years, despite its compatibility with neural networks [25, 28, 6]. The usefulness of the discrete wavelet transform in extracting specific frequency information from an image naturally fits into the realm of style transfer. Yoo *et al.* [24] were able to implement a form of wavelet pooling and unpooling in a photo-realistic style transfer model, improving upon existing models by reducing computation costs and decreasing the generation of unwanted image artifacts. Deng *et al.* [4] was able to use the stationary wavelet transform (SWT), instead of the DWT, to perform single-image super-resolution.

3. Methodology

3.1. 2D Discrete Wavelet Transform

At the core of our model lies the two-dimensional discrete wavelet transform. The DWT takes an image I and decomposes it into four filtered sub-bands, also known as wavelets, each of which is reduced to half the resolution of the original image. These four sub-bands are labeled as W_{LL} , W_{LH} , W_{HL} , and W_{HH} , each representing a unique set of information extracted from the original image. The subscript of each wavelet band refers to the order in which either low- or high-pass kernels are convolved over the original image, as seen in Eq. (1). Since convolving an image with low-pass filters smooths the image, W_{LL} is referred to as the approximation wavelet, as it heavily preserves the original image’s content while being slightly smoothed. The remaining three wavelets, W_{LH} , W_{HL} , and W_{HH} , are called detail coefficients leveraging high-pass filters to enhance directional information. These sub-bands contain vertical, horizontal, and diagonal details, respectively. In order to derive the four sub-bands for each fingerphoto I , we repeat four separate convolutions on I , each with a distinct filter kernel. For our wavelet analysis, we used the

Python PyWavelets package developed by Lee *et al.* [16]. The formulation of each kernel can be observed in Eq. (1), and example wavelet sub-bands from a sharp fingerphoto can be seen on the right-hand side of Fig. 2.

$$\begin{aligned} K_{LL} &= \frac{1}{2} \left[\frac{\sqrt{2}}{2} \frac{\sqrt{2}}{2} \right] * \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix}, \\ K_{LH} &= \frac{1}{2} \left[\frac{\sqrt{2}}{2} \frac{\sqrt{2}}{2} \right] * \begin{bmatrix} -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix}, \\ K_{HL} &= \frac{1}{2} \begin{bmatrix} -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix} * \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix}, \\ K_{HH} &= \frac{1}{2} \begin{bmatrix} -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix} * \begin{bmatrix} -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix}. \end{aligned} \quad (1)$$

This process of decomposing an image into its subsequent wavelets can be repeated to further decompose the sub-bands. Performing a Level 1 DWT on an image yields four resultant wavelets. Similarly, reapplying the process on these newly generated wavelets results in 16 sub-bands, referred to as a Level 2 DWT. This process can occur multiple times, as long as the image resolution allows it. For every level, the resolution of each wavelet halves, and the total number of wavelets increases by a factor of four. For example, an 8 x 8 image can only produce a maximum DWT of level 3, where each wavelet is a single pixel, discarding meaningful information such as mean and variance. Therefore, we must make a decision on which level wavelet decomposition to utilize in our network. Details about this choice are explained in Section 3.3.

3.2. Wavelet Style Transfer

The detail coefficients W_{LH} , W_{HL} , and W_{HH} from the DWT contain specific directional high-frequency information. Our proposed Wavelet Style Transfer (WST) takes advantage of this property to increase sharpness and deblur fingerphotos. First, we generate a set of level 3 DWT sub-bands from a blurry fingerphoto, I_b and a sharp one, I_s :

$$\{W_x^{(1)}, W_x^{(2)}, \dots, W_x^{(64)}\} = DWT(I_x). \quad (2)$$

It should be noted that these two fingerphotos do not necessarily need to share the same ID; i.e., belonging to the same subject. This step results in 128 wavelets total, where 64 are from I_b and 64 are from I_s . For each pair of wavelets $(W_b^{(i)}, W_s^{(i)})$, where i is the index of each wavelet in each set, we normalize each blurry wavelet $W_b^{(i)}$ with respect to its mean $\mu(W_b^{(i)})$ and standard deviation (STD) $\sigma(W_b^{(i)})$ to acquire the normalized wavelet $W_b^{(i)'}$. We then multiply $W_b^{(i)'}$ by $\sigma(W_s^{(i)})$ and add $\mu(W_s^{(i)})$, transferring the style

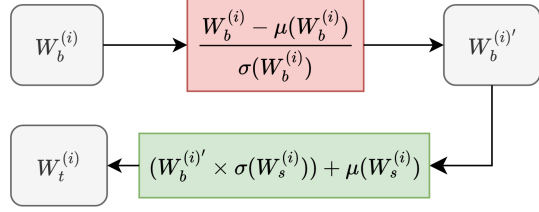


Figure 3. Our proposed Wavelet Style Transfer (WST) module, which enhances a blurry set of wavelets $W_b^{(i)}$ with a sharp set $W_s^{(i)}$, and produces a style transferred set $W_t^{(i)}$.

and high-frequency information present in I_s to I_b , and creating a style-transferred set of wavelets $W_t^{(i)}$:

$$W_t^{(i)} = WST(W_b^{(i)}, W_s^{(i)}), \text{ for } i = 1 \dots 64. \quad (3)$$

The proposed WST grants key advantages when specifically deblurring fingerphotos. From our experiments discussed in Section 3.3, it is evident that there is a significant contrast between the variance of the DWT sub-bands of a blurry and a sharp fingerphoto. Higher variance in the sub-bands is a consequence of varying gradients due to the ridges and valleys in the fingerphotos. Therefore, we can confidently hypothesize that our WST is transferring sharpness information, in the form of mean and STD, from a sharp fingerphoto to a blurry one. The proposed module can be seen in detail in Fig. 3.

3.3. DWT Levels

As stated in Sections 3.1 and 3.2, we employ a Level 3 DWT for our transfer process. This is because, depending on the selected DWT level, our deblurred images may exhibit one of two undesirable qualities upon reconstruction from the Inverse Discrete Wavelet Transform (IDWT). A Level 2 or lower DWT fails to sufficiently sharpen the output images, giving results that remain excessively blurry. Conversely, with higher levels of DWT, our deblurring results contained several artifacts. For example, decomposing the image into smaller sub-bands allows the sharp image to leak its identity information into the blurry one. Considering that we are using fingerphotos belonging to two different individuals, this causes the two separate identities to mix into a single image, thereby compromising the identity of the original image and rendering our results unusable.

Considering that all our training fingerphotos are scaled to size 256×256 , and each image contains $256^2 = 65,536$ pixels, the maximum feasible DWT level is level 8. Since $(\frac{1}{4})^8 = \frac{1}{65,536}$, and the wavelet resolution halves with each level, a level 8 DWT yields $4^8 = 65,536$ wavelets at a size of a single pixel each. Consequently, the mean of each wavelet is insignificant since each wavelet is a single pixel, and therefore, when we subtract $W_b^{(i)}$ and add $W_s^{(i)}$ dur-

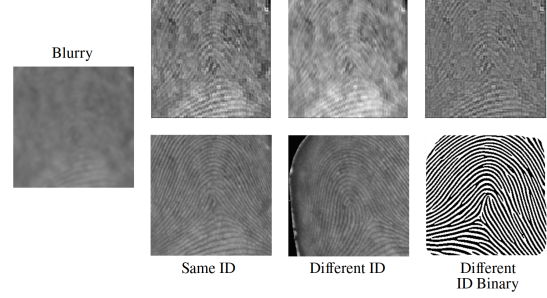


Figure 4. Sample IDWT results from a Level 3 DWT. The Blurry photo on the left is fixed, and we perform WST with three different sharp images before inverting. The first sharp image is the same ID fingerphoto as the blurry one, the second is a different ID, and the third is a binarized different ID.

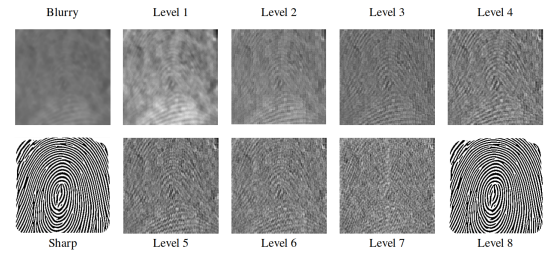


Figure 5. Examples of IDWT reconstructions from different level DWTs. The lower the level, the blurrier the reconstruction. With higher levels, artifacts become prevalent and the identities mix. Finally, at Level 8, the sharp identity fully replaces the original blurry one.

ing WST, every blurry wavelet is completely replaced by its corresponding sharp wavelet. This property can be observed at the bottom right of Fig. 5. This ID-replacement phenomenon intensifies with the DWT level from 1 to 8, but retained blurriness increases from level 8 to 1. Additionally, each level of DWT augments the total number of feature maps fed into the decoder by a factor of 4, as our wavelets are fed directly into the network. Therefore, we utilize level 3 DWT for our WST. This choice exhibited a reasonable amount of enhancement, while mitigating artifacts and maintaining the identity of the original image. Moreover, it maintains a realistic total number of wavelets. Examples of higher- and lower-level IDWT results can be observed in Fig. 5.

3.4. Binary vs. Sharp Fingerphoto

The selection of the sharp fingerphoto for WST is crucial because it directly impacts the quality of the deblurring process. We observed that DWT sub-bands with higher variance indicate a higher level of sharpness in the original fingerphoto. Hence, we advocate using an image with the highest possible variance, or sharpness, for deblurring. To this aim, we initially examined the STD of $W_b^{(i)}$ and $W_s^{(i)}$, displayed in Fig. 6. We can observe that the wavelets $W_s^{(i)}$

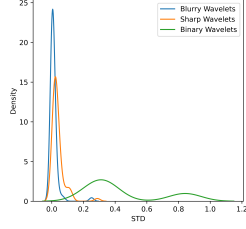


Figure 6. Density plot displaying the standard deviations of each of the sub-bands in $W_x^{(i)}$ for a sample fingerphoto, where x is either blurry (b), sharp (s), or binary (bin). The binarized fingerphoto clearly displays higher variance among its wavelets, and therefore contains higher-frequency information than the grayscale sharp or blurry version.

contain a slightly higher variance than those of $W_b^{(i)}$. This disparity is expected, and attributed to the lower frequency information present in blurry images. Based on this observation, as well as results using DWT, we devised a strategy to binarize an arbitrary sharp fingerphoto I_s into I_{bin} . This binarized fingerphoto was taken from the VeriFinger SDK.

The STD of this new I_{bin} is significantly higher than that of both I_b and I_s , indicating that theoretically it has distinctive deblurring information present. The resulting IDWT-reconstructed fingerphotos, now subjected to style-transfer with the new binarized fingerphoto, exhibit a degree of embedded sharpness but also manifest jagged artifacts. The results showing IDWT reconstructions using different style-transfer fingerphotos are provided in Fig. 4. Since the WST is adept at performing sharpness-enhancing style transfer without necessitating an encoder, we directly pass the new style-transferred wavelet set $W_t^{(i)}$ to our decoder for subsequent reconstruction and refinement.

4. Architecture

4.1. Model

Our proposed model is comprised of three main modules. Firstly, we employ a Level 3 DWT that serves as the encoder for our network. Secondly, we implement our WST module to execute style transfer and prepare the wavelets for the network. Finally, we include a decoder as our last segment. The decoder consists of a total of 9 layers. Layer 1 accepts 64 input channels, corresponding to our 64 generated wavelets, and produces 512 output channels. Subsequently, layers 2, 5, and 7 halve the total number of channels, starting from 512, and reduce them to 256, 128, and eventually back to 64 channels. All convolution kernels utilized are of size 3x3. The Rectified Linear Unit (ReLU) activation function and a 2D reflection pad are included between each layer. Additionally, layers 1, 5, and 7 incorporate an upsample operation to increase spatial size back to the original.

4.2. Loss Functions

Our network combines three loss functions for its training. They are content, style, and identity loss:

$$\mathcal{L}_{total} = \lambda_c \mathcal{L}_c + \lambda_s \mathcal{L}_s + \lambda_i \mathcal{L}_i, \quad (4)$$

where, $\mathcal{L}_c$ is content loss, $\mathcal{L}_s$ is style loss, and $\mathcal{L}_i$ is identity loss. λ_c , λ_s and λ_i are hyperparameters used to weigh each loss with respect to the others. For our final model, our loss weights were set to $\lambda_c = 100$, $\lambda_s = 10$, and $\lambda_i = 25$.

4.2.1 Content Loss

As it is beneficial in image generation tasks, we utilize L1 loss between the original sharp fingerphoto I_s and its deblurred version I_{deblur} , over a batch of size n , in order to retain as much identity information as we can upon deblurring. This helps remove arbitrarily generated minutiae that do not reflect I_s and preserve the original content image as much as possible while simultaneously sharpening edges, as shown here:

$$\mathcal{L}_c = \frac{1}{n} \sum_{i=1}^n |I_{deblur}^{(n)} - I_s^{(n)}|. \quad (5)$$

4.2.2 Style Loss

We incorporate style loss using the generated feature maps of a pretrained VGG19 encoder. To better retain both local and global style features from our deblurred images, we use all 5 activation layers of the VGG encoder for our feature map generation. Inspired by [10], we then take the feature representations of both I_s and I_{deblur} , and calculate the difference between their mean and STD:

$$\begin{aligned} \mathcal{L}_s = & \sum_{\phi=1}^5 \|\mu(V_\phi(I_{deblur})) - \mu(V_\phi(I_s))\|_2 \\ & + \sum_{\phi=1}^5 \|\sigma(V_\phi(I_{deblur})) - \sigma(V_\phi(I_s))\|_2. \end{aligned} \quad (6)$$

Here, $V_\phi(\cdot)$ represents the feature maps generated by the corresponding ϕ^{th} activation layer of a pretrained VGG19 encoder.

4.2.3 Identity Loss

For our Identity Loss, we incorporated a separate pre-trained ResNet-18 verifier to preserve ID information throughout the deblurring process. This verifier is trained to match fingerphotos in the embedded domain and extracts multi-level features of a given fingerphoto. We take the

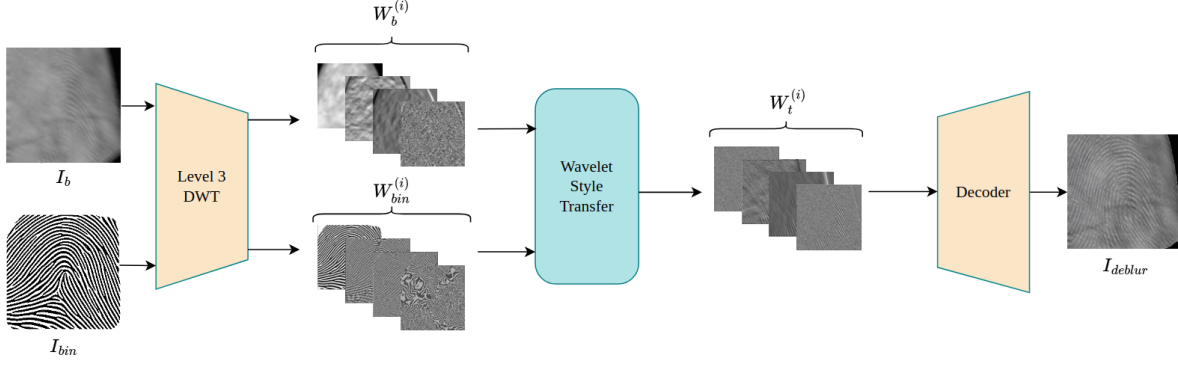


Figure 7. Our proposed deblurring model. Two input fingerphotos I_b and I_{bin} are first passed into the DWT, producing two sets of wavelets $W_b^{(i)}$ and $W_{bin}^{(i)}$. The resulting wavelets are merged, transferring the style from I_{bin} onto I_b . We then treat this combined set of wavelets $W_t^{(i)}$ as feature maps for our decoder, which combines them back into a deblurred fingerphoto I_{deblur} .

Mean Squared Error (MSE) loss between the embeddings, and then between the features of I_s and I_{deblur} :

$$\mathcal{L}_s = \|E(I_{deblur}) - E(I_s)\|_2 + \sum_{i=1}^3 \|F_i(I_{deblur}) - F_i(I_s)\|_2, \quad (7)$$

where, E represents the embeddings from the last layer of the verifier, and F_i represents the generated features at the i^{th} intermediate layer within the verifier. This helps ensure that the features of I_{deblur} , or the minutiae, closely match those of I_s , subsequently maintaining identity through transfer.

5. Experiments

5.1. Training

For our training, we trained the model for a total of 240 epochs. We use the Adam optimizer, with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and an initial learning rate of 0.0001. Our model was trained on three NVIDIA GeForce GTX 1080 Ti.

For our training dataset, we used the WVU Multimodal Fingerphoto Dataset [2]. This dataset consists of 3,542 fingerphotos, all of which have a resolution 256×256 . We blurred each fingerphoto during training, using one of three equally likely blurring methods: Gaussian, Motion, or Average. It should be noted that the kernel size for all three blur types is given as $k = 6\sigma - 1$ where σ is an integer from range [3, 6]. We chose Gaussian Blur's σ to be in range [3, 6], and Average Blur to use a kernel size of $\frac{k}{2}$, as a kernel size of k produced images that were unrealistically blurry to contain any relevant identifying information.

5.2. Evaluation

Since our method integrates both style transfer and deblurring, we compare our method with other style trans-

fer and deblurring architectures. To measure our performance, we fed our deblurred results to both Neorutechnology's Verifinger and Innovatrics' IDKit to obtain evaluations from multiple matchers. Our primary evaluation metrics comprise Receiver Operating Characteristic (ROC) curves, and are Area Under the Curve (AUC), and Equal Error Rate (EER). ROC assesses the model's ability to classify deblurred fingerphotos into their sharp counterparts (genuine pairs) as well as its capacity to distinguish imposter pairs, where a sharp fingerphoto belongs to a different identity. The AUC can be interpreted as a measure of the accuracy of the model. Additionally, we use the EER metric, which indicates the point at which the model experiences equal occurrences of False Positives and False Negatives. In addition to ROC curves, we also used NFIQ2 V2.2, from NIST [1] to evaluate the quality of our generated fingerphotos compared to their originals. By combining both quality and verifier evaluations, we can ensure that our deblurred fingerphotos retain their identity and increase sharpness.

6. Results

We evaluated our model on multiple different test sets with variations in blurring. Three test sets were used, the Multimodal Dataset [2], from West Virginia University, the Touch and Touchless Fingerprint Dataset [3] from the Indian Institute of Technology, Bombay, and the RidgeBase Benchmark Dataset [12], from the University of Buffalo. These test sets contained 309, 800, and 2,229 images, respectively. For testing, we created genuine pairs using different instances of the same finger, and for impostor pairs, we randomly selected a sample from a different subject.

As depicted in Fig. 8 and Table. 1, our FDWST outperforms other style transfer and deblurring methods when deblurring fingerphotos. Due to the absence of proper blurry-sharp pairs for testing, we initially blurred each dataset and then passed these images through each model to obtain our

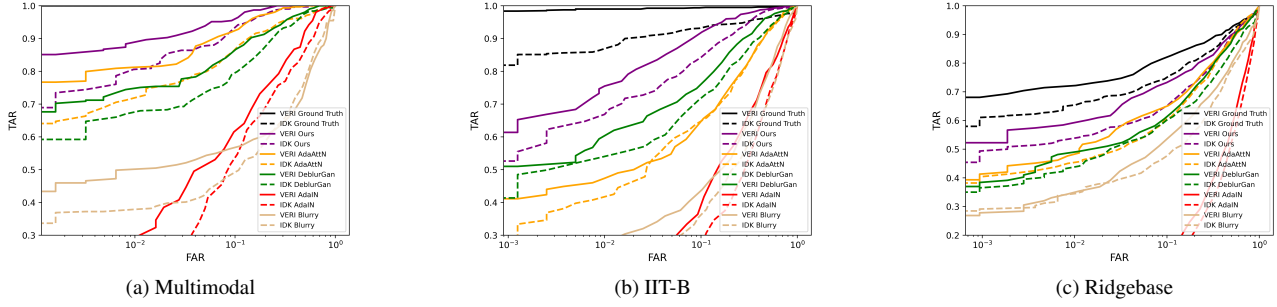


Figure 8. ROC curves on three Datasets: (a) WVU Multimodal, (b) IIT-Bombay, (c) Ridgebase. Solid lines correspond to VeriFinger results, and dashed lines represent IDKit results. Best seen in color. Zoom in for better view.

Model	IDKit			VeriFinger		
	Multimodal	IIT-B	Ridgebase	Multimodal	IIT-B	Ridgebase
	AUC/EER	AUC/EER	AUC/EER	AUC/EER	AUC/EER	AUC/EER
Ground Truth	1.0000/0.0000	0.9583/0.0725	0.8835/0.1931	1.0000/0.0000	0.9958/0.0100	0.9075/0.1548
Blurry	0.7514/0.3334	0.6821/0.3746	0.7336/0.3285	0.7565/0.3407	0.7477/0.3379	0.7647/0.3057
AdaIN	0.8106/0.2549	0.6559/0.3868	0.5095/0.4876	0.8575/0.2312	0.7347/0.3227	0.5556/0.4654
AdaAttN	0.9506/0.1154	0.8622/0.2271	0.8366/0.2593	0.9812/0.0844	0.8669/0.2271	0.8442/0.2389
DeblurGAN	0.9315/0.1571	0.8888/0.1939	0.8039/0.2779	0.9548/0.1213	0.9185/0.1698	0.8334/0.2521
FDWST	0.9795/0.0852	0.9483/0.1297	0.8720/0.2074	0.9907/0.0545	0.9703/0.0908	0.8720/0.2074

Table 1. Quantitative results comparing our FDWST against other models, over all three datasets WVU Multimodal, IIT-Bombay, and Ridgebase, and from two verifiers, VeriFinger and IDKit.

Dataset	Before	After
WVU Multimodal	0.5954	23.2168
IIT-Bombay	12.7204	39.4500
RidgeBase	15.2445	25.6419

Table 2. Quality Scores of different datasets before and after being passed through our proposed model. Each dataset shows a substantial increase in quality after the deblurring process.

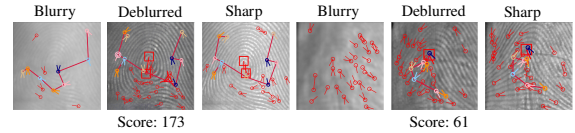


Figure 9. Naturally blurred fingerphotos, with their corresponding minutiae maps from VeriFinger SDK. The scores at the bottom of the figure correspond to deblurred and sharp pair. VeriFinger failed to generate matching score for blurry and deblurred pair.

results. The results in Figs. 8 and 10, which indicate that the FDWST achieves the highest matching accuracy (AUC), coupled with a significant increase in image quality upon deblurring as shown in Table. 2, reaffirms that our model deblurs fingerphotos in such a way that retains their original identity while also generating realistic minutiae. In addition, we also evaluated our model on available naturally blurred samples. These results are depicted in Fig. 9. It can be observed that the original blurry images contain numerous incorrect minutiae, or none at all, whereas our model is able to accurately reconstruct relevant minutiae.

7. Ablation Study

For our ablation study, we examined the impact of removing different loss terms from the network, as well as comparing performance by eliminating the decoder and em-

ploying the IDWT operation for reconstruction.

7.1. Stand-alone IDWT

Initially we experimented with the Inverse Discrete Wavelet Transform (IDWT). Since DWT is a linear transformation, any set of wavelets $W_I^{(i)}$ from image I can be used to perfectly reconstruct the original image using IDWT. First, we performed the WST operation on $(W_b^{(i)}, W_s^{(i)})$ to achieve a style-transferred set of wavelets $W_t^{(i)}$, and then fed these new wavelets to the IDWT for reconstruction. We noticed that this linear process helps to deblur images. Furthermore, any sharp I_s could be used to enhance blurred I_b indicating that sharpness information is being transferred between the two images. To examine the effectiveness of our WST operation, we fed resultant images from the

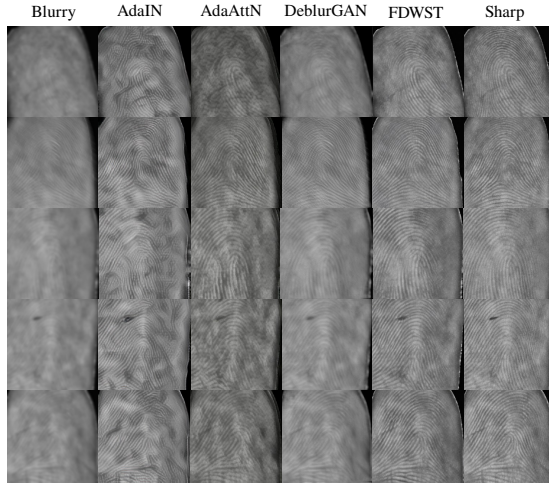


Figure 10. Sample fingerphotos from each model, as well as the original sharp and blurry versions.

IDWT operation directly to VeriFinger, and attempted to match them to their sharp counterparts. We first investigated the performance using a grayscale fingerphoto for the style transfer process, and then swapped it out for a binarized image. As seen in Fig. 11, the WST process followed by IDWT, notably without a decoder network, performs well.

As illustrated in Table 3, utilizing a binarized fingerphoto as opposed to a sharp one yields a significant improvement in performance. Our binary fingerphoto led us to a 2.3% increase in AUC, as well as a 50% drop in EER, compared to reconstructing with a sharp grayscale fingerphoto. This outcome influenced our decision to use a binarized fingerphoto to perform style transfer in our final proposed model. The IDWT operation surpasses our models with certain loss terms removed, exhibiting better performance than both our content loss and style loss removed models. With just Identity Loss removed, the binary IDWT operation performed nearly identically to our model, with changes of only 0.0016 and 0.0008 in AUC and EER, respectively. We believe that these results support the claim that our WST operation is transferring high-frequency and sharpness information from a binarized fingerphoto to a blurry one.

7.2. Removed Loss Terms

The removal of any loss term from our network reduced performance, with some having more significant impacts than others. In particular, eliminating identity loss did not greatly affect performance, which was anticipated given that identity loss plays a smaller role than either content or style loss. The purpose of identity constraint in our network is to catch and correct minor minutiae discrepancies between deblurred and sharp fingerphotos, so its removal does

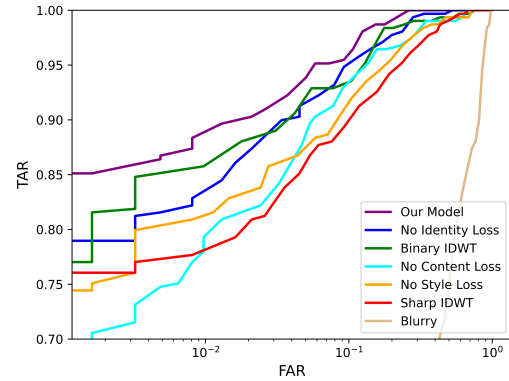


Figure 11. ROC curve results from different versions of the model obtained using VeriFinger SDK.

Model	AUC	EER
Our Deblurring Model	0.9907	0.0545
Model Without Identity Loss	0.9847	0.0719
Binary IDWT	0.9831	0.0711
Model Without Content Loss	0.9759	0.0819
Model Without Style Loss	0.9736	0.0916
Sharp IDWT	0.9680	0.1005
Blurry Test Set	0.7565	0.3407

Table 3. Comparisons of different versions of the model by eliminating certain loss terms. Metrics obtained from VeriFinger SDK.

not affect the network as broadly as the other loss terms.

However, without content loss, our network produced visually appealing images that lost a significant portion of their identity. This underscores the critical role of content loss in preserving the identity of the generated fingerphotos. Furthermore, removing style loss from the network resulted in fingerphotos that did not achieve a desirable level of deblurring in terms of overall structure and sharpness. This observation suggests that style loss, since it is calculated on the basis of mean and STD, measures the level of sharpness transferred during WST. A comprehensive comparison of all loss-removed models can be seen in Fig. 11 and Table 3.

8. Conclusion

In summary, we proposed a fingerphoto deblurring architecture FDWST that utilizes techniques from both style transfer and the discrete wavelet transform. We demonstrated that our model not only improves fingerphoto quality, but also improves matching performance across multiple verifiers. Based on these findings, we assert that our Wavelet Style Transfer is proficient in transferring sharpness information and holds significant potential for various future applications.

References

- [1] NFIQ2. Available at <https://www.nist.gov/services-resources/software/nfiq-2>. Accessed: March 26, 2024.
- [2] WVU Multimodal Database. Available at <https://biic.wvu.edu/data-sets/multimodal-dataset>. Accessed: March 17, 2024.
- [3] P. Birajadar, M. Haria, P. Kulkarni, S. Gupta, P. Joshi, B. Singh, and V. Gadre. Towards smartphone-based touchless fingerprint recognition. *Sādhanā*, 44, 07 2019.
- [4] X. Deng, R. Yang, M. Xu, and P. L. Dragotti. Wavelet domain style transfer for an effective perception-distortion tradeoff in single image super-resolution. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 3076–3085, 2019.
- [5] Y. Deng, F. Tang, W. Dong, C. Ma, X. Pan, L. Wang, and C. Xu. Stytr2: Image style transfer with transformers. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11326–11336, 2022.
- [6] Y. Duan, F. Liu, L. Jiao, P. Zhao, and L. Zhang. Sar image segmentation based on convolutional-wavelet neural network and markov random field. *Pattern Recognition*, 64:255–267, 2017.
- [7] L. A. Gatys, A. S. Ecker, and M. Bethge. A neural algorithm of artistic style, 2015.
- [8] M. Hradiš, J. Kotera, P. Zemčík, and F. Šroubek. Convolutional neural networks for direct text deblurring. In *Proceedings of BMVC*, volume 10, 2015.
- [9] H. Huang, H. Wang, W. Luo, L. Ma, W. Jiang, X. Zhu, Z. Li, and W. Liu. Real-time neural style transfer for videos. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 783–791, 2017.
- [10] X. Huang and S. Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In *Proceedings of the IEEE international conference on computer vision*, pages 1501–1510, 2017.
- [11] Innovatrics. Innovatrics abis: Biometrics, tailored. *IDKit SDK*. Available at <https://www.innovatrics.com/innovatrics-abis/>. Accessed: March 15, 2024.
- [12] B. Jawade, D. Mohan, S. Setlur, N. Ratha, and V. Govindaraju. *RidgeBase: A Cross-Sensor Multi-Finger Contactless Fingerprint Dataset*. 2022 IEEE International Joint Conference on Biometrics (IJCB), 2022.
- [13] A. S. Joshi, A. Dabouei, J. Dawson, and N. M. Nasrabadi. Fingerphoto deblurring using attention-guided multi-stage gan. *IEEE Access*, 2023.
- [14] D. Kadish, S. Risi, and A. S. Løvlie. Improving object detection in art images using only style transfer. In *2021 international joint conference on neural networks (IJCNN)*, pages 1–8. IEEE, 2021.
- [15] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas. Deblurgan: Blind motion deblurring using conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8183–8192, 2018.
- [16] G. R. Lee, R. Gommers, F. Waselewski, K. Wohlfahrt, and A. O’Leary. Pywavelets: A python package for wavelet analysis. *Journal of Open Source Software*, 4(36):1237, 2019.
- [17] S. Liu, T. Lin, D. He, F. Li, M. Wang, X. Li, Z. Sun, Q. Li, and E. Ding. Adaattn: Revisit attention mechanism in arbitrary neural style transfer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6649–6658, 2021.
- [18] F. Luan, S. Paris, E. Shechtman, and K. Bala. Deep photo style transfer. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4990–4998, 2017.
- [19] S. Nah, T. Hyun Kim, and K. Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3883–3891, 2017.
- [20] Neurotechnology. *VeriFinger SDK*. Available at <https://www.neurotechnology.com/verifinger.html>. Accessed: March 15, 2024.
- [21] D. Y. Park and K. H. Lee. Arbitrary style transfer with style-attentional networks. In *proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5880–5888, 2019.
- [22] X. Tao, H. Gao, X. Shen, J. Wang, and J. Jia. Scale-recurrent network for deep image deblurring. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8174–8182, 2018.
- [23] L. Xu, J. Ren, C. Liu, and J. Jia. Deep convolutional neural network for image deconvolution. *Advances in Neural Information Processing Systems*, 2:1790–1798, 01 2014.
- [24] J. Yoo, Y. Uh, S. Chun, B. Kang, and J.-W. Ha. Photorealistic style transfer via wavelet transforms. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9036–9045, 2019.
- [25] J. Zhang, G. G. Walter, Y. Miao, and W. N. W. Lee. Wavelet neural networks for function learning. *IEEE transactions on Signal Processing*, 43(6):1485–1497, 1995.
- [26] K. Zhang, W. Luo, Y. Zhong, L. Ma, B. Stenger, W. Liu, and H. Li. Deblurring by realistic blurring. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2737–2746, 2020.
- [27] X. Zhang, R. Jiang, T. Wang, and J. Wang. Recursive neural network for video deblurring. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(8):3025–3036, 2020.
- [28] B. Zhou, A. Shi, F. Cai, and Y. Zhang. Wavelet neural networks for nonlinear time series analysis. In *Advances in Neural Networks-ISNN 2004: International Symposium on Neural Networks, Dalian, China, August 19-21, 2004, Proceedings, Part II 1*, pages 430–435. Springer, 2004.