NIST Grant/Contractor Report
NIST GCR 23-040

An Investigation of Applications of Neural Style Transfer to Forensic Footwear Comparison

Gregory J. Stock

This publication is available free of charge from:
https://doi.org/10.6028/NIST.GCR.23-040

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

NIST Grant/Contractor Report
NIST GCR 23-040

An Investigation of Applications of Neural Style Transfer to Forensic Footwear Comparison

Gregory J. Stock
Associate, Statistical Engineering Division
Information Technology Laboratory

Georgetown University

This publication is available free of charge from:
https://doi.org/10.6028/NIST.GCR.23-040

February 2023

U.S. Department of Commerce
Gina M. Raimondo, Secretary

National Institute of Standards and Technology
Laurie E. Locascio, NIST Director and Under Secretary of Commerce for Standards and Technology

This publication was produced as part of the National Institute of Standards and Technology Professional Research Experience Program (PREP). The contents of this publication do not necessarily reflect the views or policies of the National Institute of Standards and Technology or the US Government.

How to cite this NIST Technical Series Publication:
Gregory J. Stock (2023) An Investigation of Applications of Neural Style Transfer to Forensic Footwear Comparison. (National Institute of Standards and Technology, Gaithersburg, MD), NIST GCR 23-040. https://doi.org/10.6028/NIST.GCR.23-040

Contact Information
gregory.stock@nist.gov

NIST GCR 23-040
February 2023

Abstract

Neural Style Transfer (NST) is an application of neural networks that allows one to create an image that has the content of one image and the style of another image. For example, NST can be used to create an image of a cat in the style of Van Gogh's *Starry Night*. This study asks two distinct questions. Can we use NST to manufacture pseudo crime scene images of shoe impressions by transferring the style of a crime scene image to a clean impression of a shoe? The idea is that the known, clean image will be transformed to take on the characteristics of a crime scene image, thereby providing a mechanism for researchers to generate pseudo crime scene images that can be used, for example, in training impression recognition algorithms. This is useful because there is a shortage of ground truth known crime scene-like data. We report that our NST procedure does not produce extremely convincing pseudo impressions but the results are nonetheless promising. The second question asks if we can improve crime scene/suspect shoe impression comparison by using NST as a preprocessing, sharpening step. The idea is that the "style" of well-defined edges of the clean impressions will be transferred to the crime scene image. We have a lineup of clean shoe impressions containing the impression of the shoe that created the crime scene image and several other impressions of shoes of the same make, model, and size. We compare the crime scene image to each of the shoes in the lineup using an algorithm that computes similarity scores. The matching shoe should give the highest score with high probability. Next, we combine the crime scene image with an aggregate image of all the clean shoe impressions in the lineup using NST. Then we repeat the comparison process with the NST image instead of the original crime scene image. We observe that, in half of the cases we studied, when we compare the NST image to each of the shoes in the lineup, we can more readily distinguish between the matching shoe and the non-matching shoes than when we do the comparisons with the original crime scene image.

Keywords

Convolutional neural network; Footwear comparison; Forensic analysis; Neural style transfer, Similarity metrics.

Table of Contents

List of Tables

List of Figures

Acknowledgments

The author wishes to thank Steven Lund for sharing his flexible image alignment technique and the images it produced that were used in this paper. The author also wishes to thank Hari Iyer for sharing his comments on a preliminary draft of this paper.

1. Introduction

Forensic footwear analysis in the United States is conducted by expert human analysts. One of the goals of analysis is to determine if a test impression, K, of a known shoe of interest, matches the image of a crime scene impression, Q. A shortcoming of the current process is its subjectivity. There is an ongoing effort [2] to add objective, quantitative measures to the process using computer algorithms. One type of algorithm centers around similarity measures that answer questions like, "how similar or dissimilar is a crime scene impression to a test impression." A second type is neural networks. Neural networks can be trained to determine when a K matches a Q. These networks need to be trained with many crime scene images. Finding a source of training images is a challenge. It would be useful to have a means of generating artificial crime scene images to serve as training images. We explore the utility of the deep learning method known as neural style transfer (NST) as a tool in both computing similarity scores and in producing training images.

Neural Style Transfer is a deep learning algorithm that combines the content of a target image with the style of a reference image. See, for example, Fig. 1 in which the content of an image of building fronts is being combined with the style of Van Gogh's *Starry Night*. It is accomplished with a convolutional neural network that creates an image x such that the following loss function is minimized,

$$L = a\,|C(p) - C(x)| + b\,|S(q) - S(x)| \tag{1}$$

where, $C(\cdot)$ is a function of an image representing its content, $S(\cdot)$ is a function of an image representing its style, p is the content image, q is the style image, x is the output image, a and b are weights. For definitions of $C(\cdot)$ and $S(\cdot)$ and a more complete understanding of the algorithm, see Ref. [3]. Minimizing L means that the style of the output image is close to the style of the reference image and the content of the output image is close to the content of the target image.

Fig. 1. A demonstration of neural style transfer. From Deep Learning with R, [1]

In this work we explore two applications of neural style transfer to footwear comparison that answer two distinct questions: 1) Can we use neural style transfer to create artificial crime scene images by combining the content of a K with the style of a Q ? (See Figure 2)

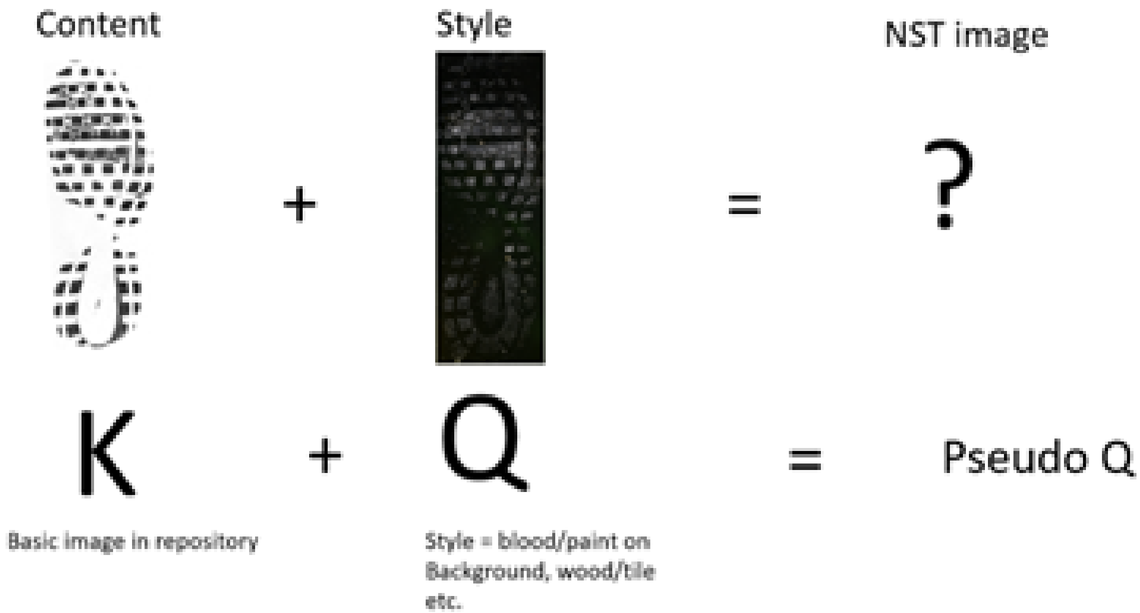


Fig. 2. A schematic of the first application.

2) Can we transfer the "sharpness" of a K to a crime scene image Q before comparing it to a test impression to improve the comparison process? (See Figure 3)

To answer these questions, we applied neural style transfer to staged crime scene impressions and test impressions created by NIST researchers from 3 pairs of Nike Dual Fusion sneakers.[1] We use the neural style algorithm in Deep Learning with R [1] to create our NST images.

2. Staging the Crime Scene Impressions

The images used in this study were created by Steve Lund[2], Hariharan Iyer, and Gunay Dogan, who are researchers at the National Institute of Standards and Technology, Martin Herman (a guest researcher at NIST), Vighnesh Hegde, Gautham Venkatasubramanian and Sarah Hood (former guest researchers at NIST), and Mike Gorn (a footwear impression examiner from the FBI). The purpose of creating the images was to have several sets of

[1]Certain commercial equipment, instruments, or materials are identified in this paper to foster understanding. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.
[2]These images are available to interested readers by contacting steven.lund@nist.gov.

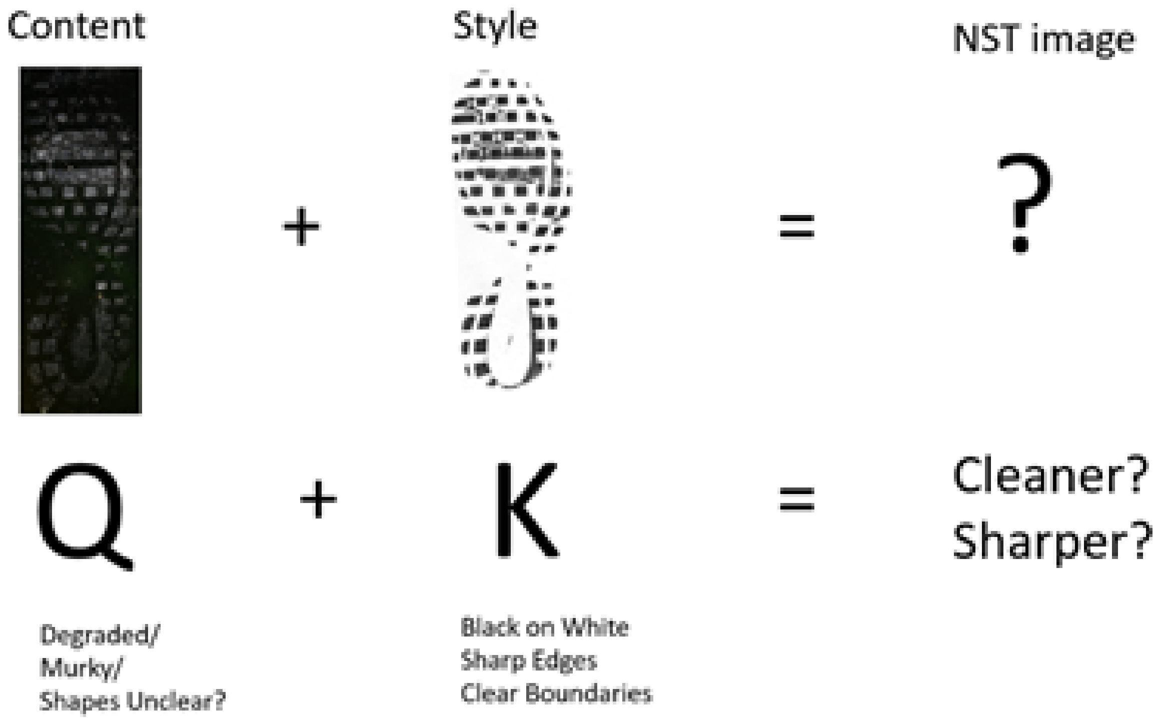


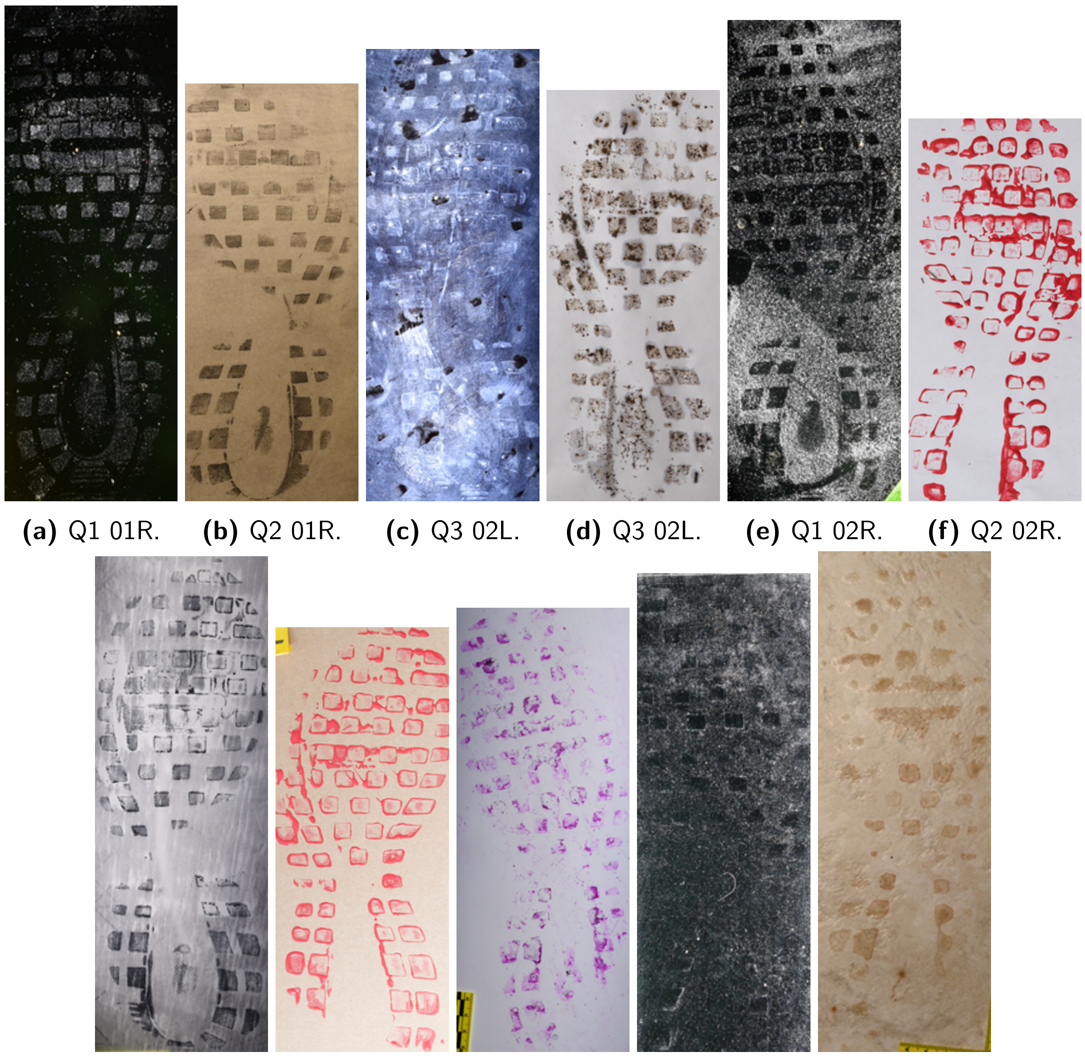
Fig. 3. A schematic of the second application.

images, each consisting of a staged crime scene impression, a test impression of the shoe that created it and test impressions of several close non-matches. A close non-match is a shoe of the same make, model and size as the shoe that created the crime scene impression.

The images were created from 3 pairs of sneakers of the same make, model, and size. To create the test impressions, a researcher rolled each of the six shoes from heel to toe over a 2D EverOS scanner while wearing the shoe, thus capturing the entire outsole of the shoe. The impression process was performed 5 times for each shoe, so that there was a total of 30 impressions. The images were labeled with the following scheme: Shoe Pair Number, Left or Right Shoe, Impression Number. For example, 02L 03 refers to the third impression of the left-hand shoe of the second pair. The right shoe images of each pair were then flipped horizontally so that all the impressions appeared to be left-handed.

Eleven crime scene impressions were created from the six shoes. These included wet and dry impressions on surfaces including paper, tile, counter top, wood, glass, and cardboard. Oblique lighting, electrostatic lifts, and gel lifts were used when appropriate to collect impressions. Each crime scene impression, Q, was labeled with the shoe that created it, for example 03L, the left-hand shoe of the third pair.

Now each Q is paired with the 30 test impression images, 5 of which come from the match-



(a) Q1 01R. **(b)** Q2 01R. **(c)** Q3 02L. **(d)** Q3 02L. **(e)** Q1 02R. **(f)** Q2 02R.

(g) Q3 03L. **(h)** Q3 03L. **(i)** Q3 03R. **(j)** Q3 03R. **(k)** Q3 03R.

Fig. 4. Eleven staged crime scene impressions.

ing shoe that created it and 25 from the close non-matches, completing the image creation by the aforementioned group of researchers.

The test impressions were further processed as follows. For every Q, each of its 30 test impressions was aligned to it by an alignment procedure developed at NIST by Steve Lund. The alignment goes beyond the rotation and translation of a rigid alignment by dividing the shoe image into patches that are allowed to rotate and transform independently of each other. This helps to achieve a more realistic alignment to an image that was created by rolling, and thus bending and stretching the sole.

For each of the 30 test impression images, for every Q, a binary version and a signed edge distance version of the test impressions were created. In the binary version, the pixels above a threshold are assigned the value 255 (white) while pixels below the threshold are assigned the value 0 (black). The signed edge distance is computed by assigning pixel values 0 to 50, such that pixels that constitute an edge of a feature are assigned the value 50, pixels that are 1 pixel away from an edge are assigned the value 49, ... , pixels that are 50 or more pixels away are assigned the value 0. Finally for each Q, its 30 flex-aligned images were averaged into an aggregate flex image, the 30 binary images were averaged into an aggregate binary image and the 30 signed edge distance images were averaged into an aggregate signed edge distance image.

(a) Flex alligned. **(b)** Binary. **(c)** Signed edge distance.

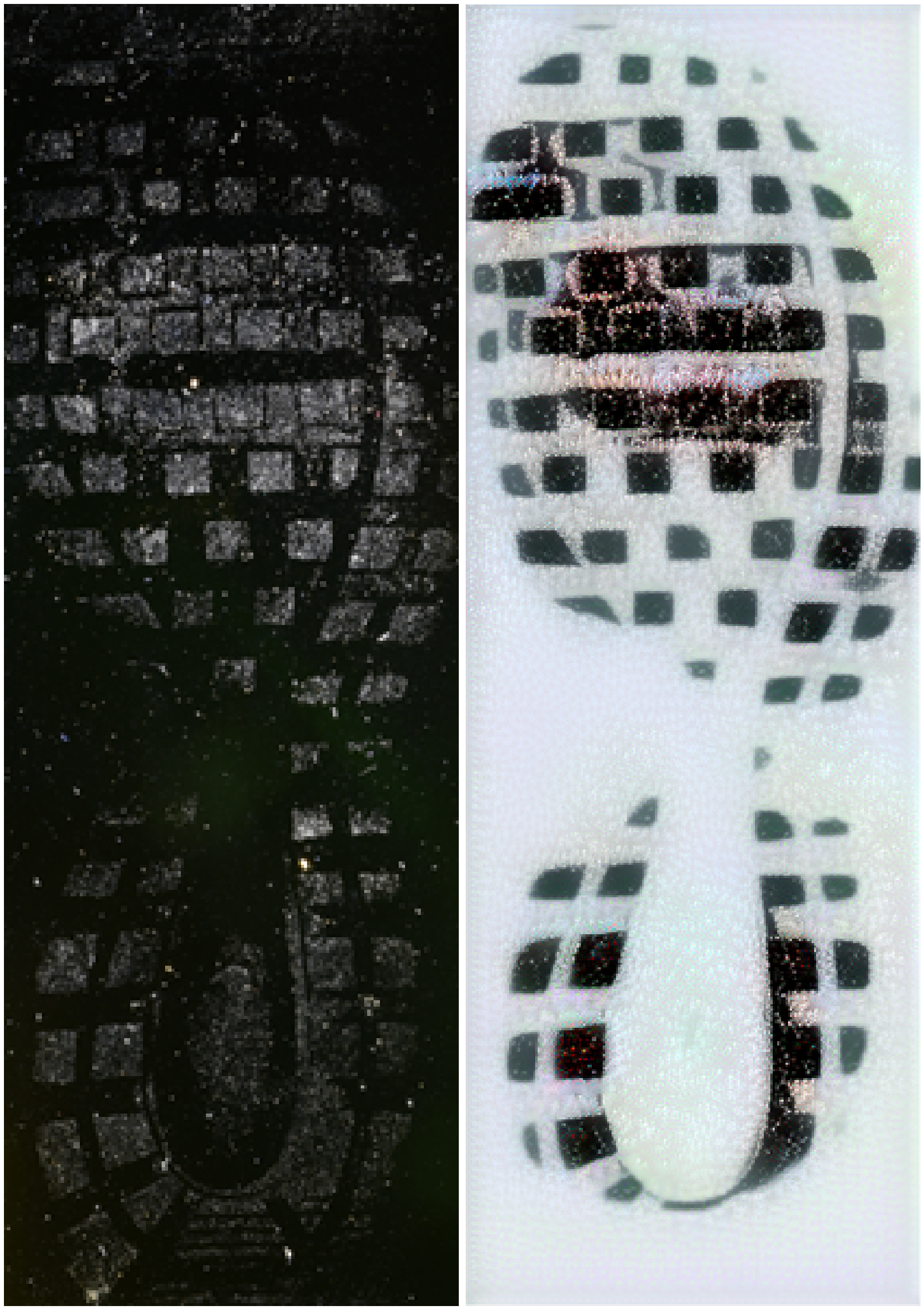
Fig. 5. Three versions of each test impression.

3. Creating Artificial Qs

Here we investigate the usefulness of NST in creating artificial crime scene images. If successful, we would have a source of ground truth known images, that is, a source of crime scene images for which the shoes that generated them are known. This could be useful in training neural networks to identify matching impressions.

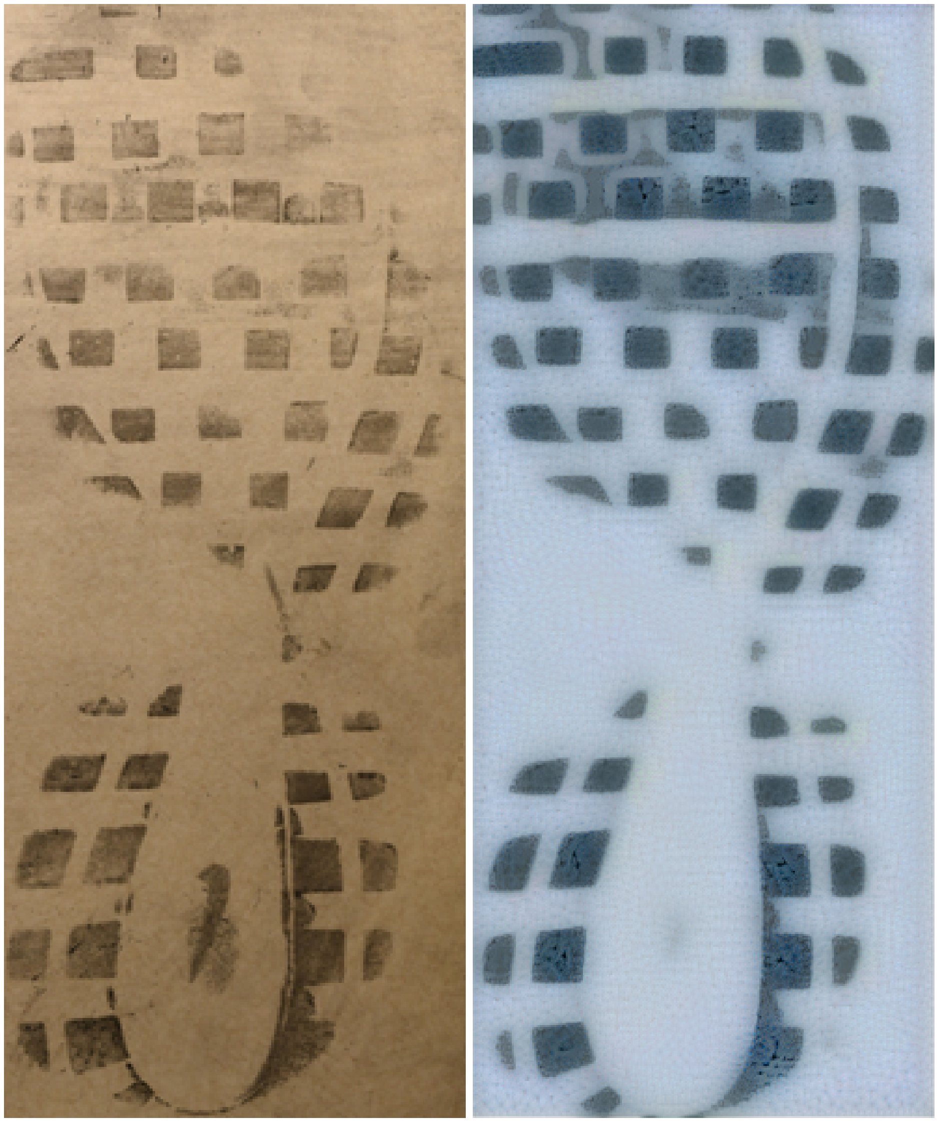
We use NST to combine the style of a Q with the content of a K. Our desire is that the texture of the crime scene image will be transferred to the clean test impression K. We combine each of the eleven crime scene images with the first test impression of the generating shoe. For example we combine the style of the Q1 image with the content of the 01R-01 test impression. We must choose weights for the style and content components. So we create five different NSTs with weights $(a,b) = \{(.925,.1),(.425,.6),(.125,.9),(.025,1.0),(10^2,10^8)\}$. We have included the last weight setting to greatly emphasize style over content.

We display the $(10^2,10^8)$ weighted image for all eleven crime scene images. How well do the visual textures, for instance, the appearance of blood, transfer to the K impression? We provide a qualitative visual assessment of the success of the style transfer. In future work, others may attempt to find a quantitative measure of success. Figures 6 through 16 show side-by-side comparisons of the original Q image and its pseudo Q NST image. We observe that for Pseudo Q1, some of the speckling of the Q1 image has transferred to the region of the shoe by the ball of the foot and in the heel; the brown speckling found in the bottom left of the Q2 image has transferred to the Pseudo Q2 image; the style of the Q3 image does not appear to have transferred to the Pseudo Q3 image; the speckling pattern of the Q4 image appears to have transferred to the Pseudo Q4 image; the speckling in the background of the Q5 image appears to have transferred to the Pseudo Q5 image; the accumulation of paint around the tread of the outsole in image Q6 has transferred to the Pseudo Q6 image (just not in red); the speckling around the tread and the vertical streaking of the Q7 image has transferred to the Pseudo Q7 image; the accumulation of paint around the tread of the outsole in image Q8 has transferred to the Pseudo Q8 image; the light diagonal streaking of image Q8 appears to have transferred to the Pseudo Q9 image; the marble texture of image Q10 does not appear to have transferred to the Pseudo Q10 image; the texture of the depressions in the tile do not appear to have transferred to the Pseudo Q11 image.



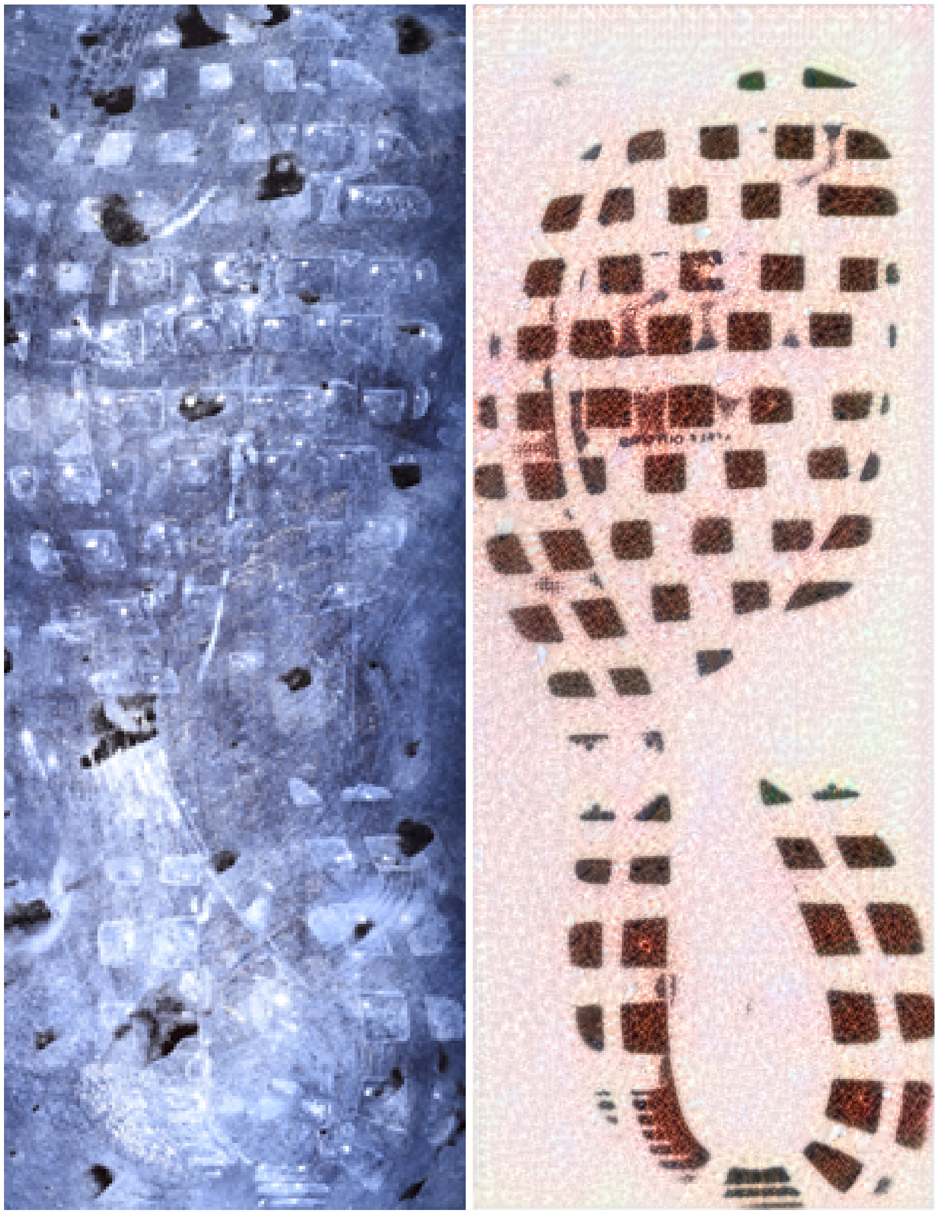
(a) Q1 01R. **(b)** Pseudo Q1.

Fig. 6. A NST of the style of Q1 to the content of 01R 01.



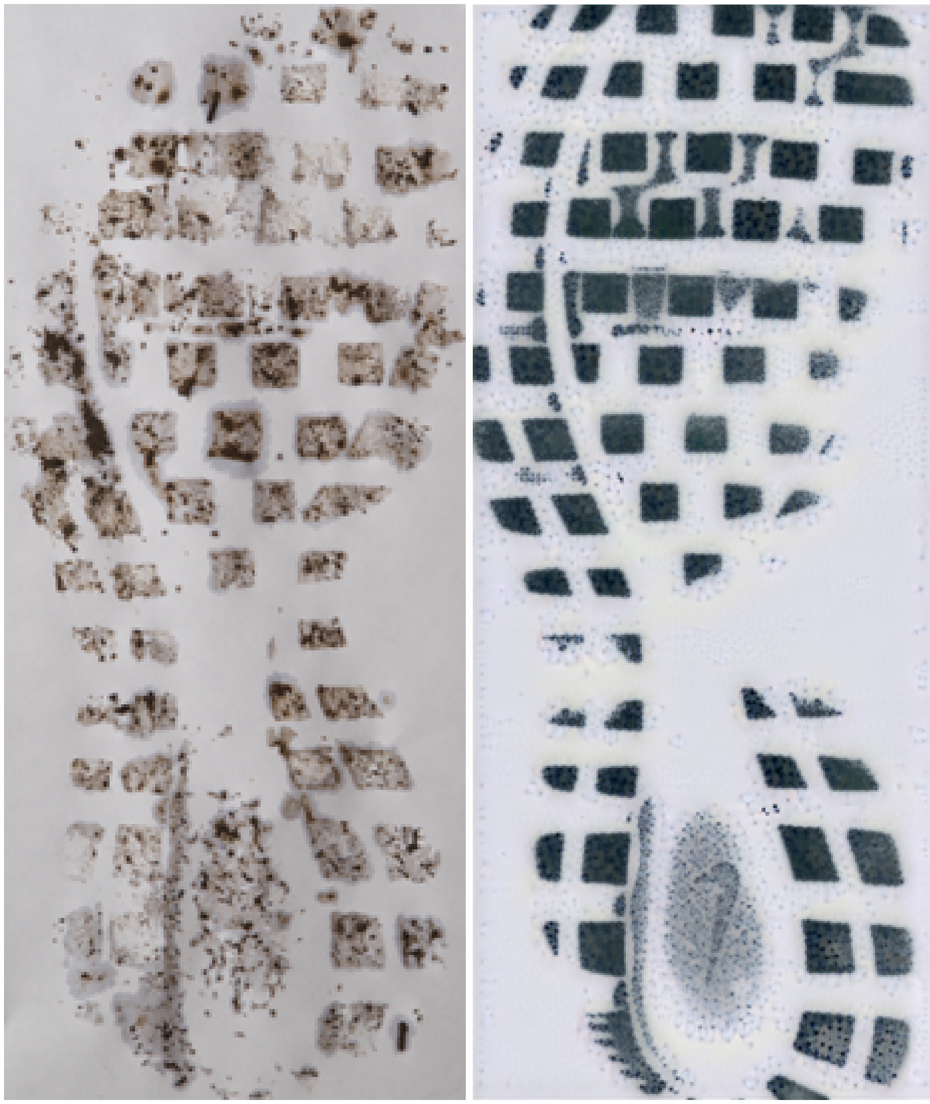
(a) Q2 01R. **(b)** Pseudo Q2.

Fig. 7. A NST of the style of Q2 to the content of 01R 01.



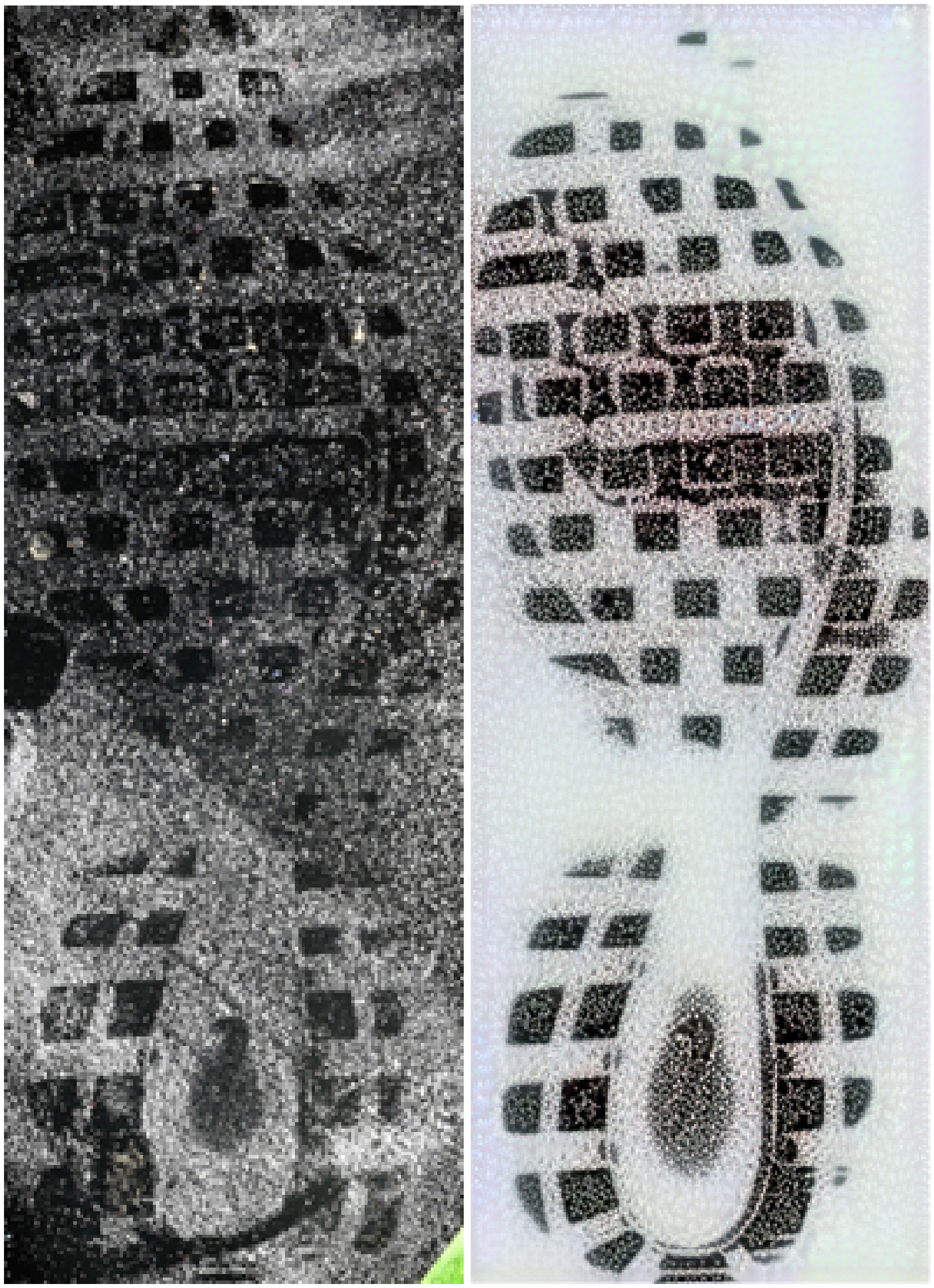
(a) Q3 02L. **(b)** Pseudo Q3.

Fig. 8. A NST of the style of Q3 to the content of 02L 01.



(a) Q4 02L. **(b)** Pseudo Q4.

Fig. 9. A NST of the style of Q4 to the content of 02L 01.

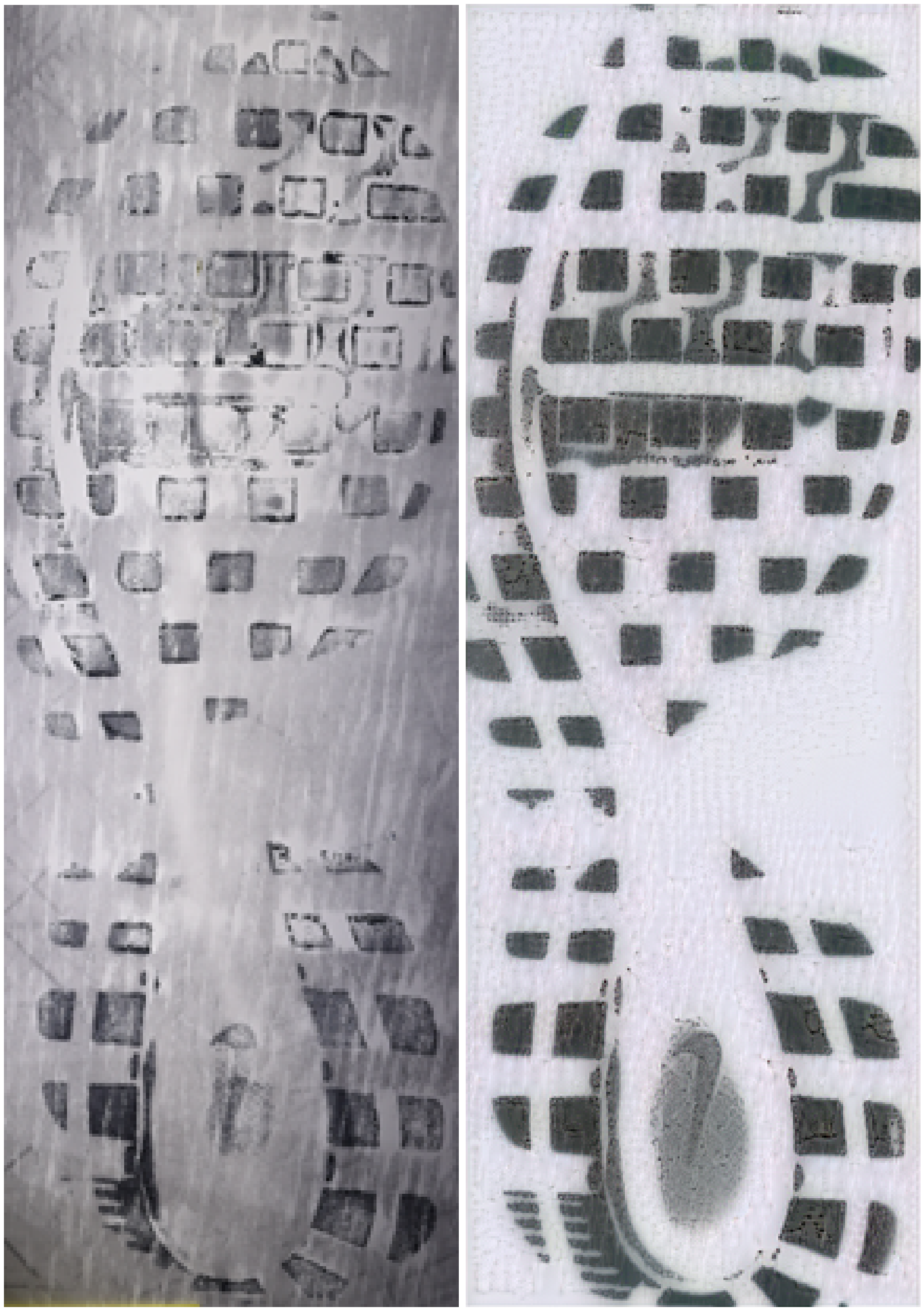


(a) Q5 02R. **(b)** Pseudo Q5.

Fig. 10. A NST of the style of Q5 to the content of 02R 01.

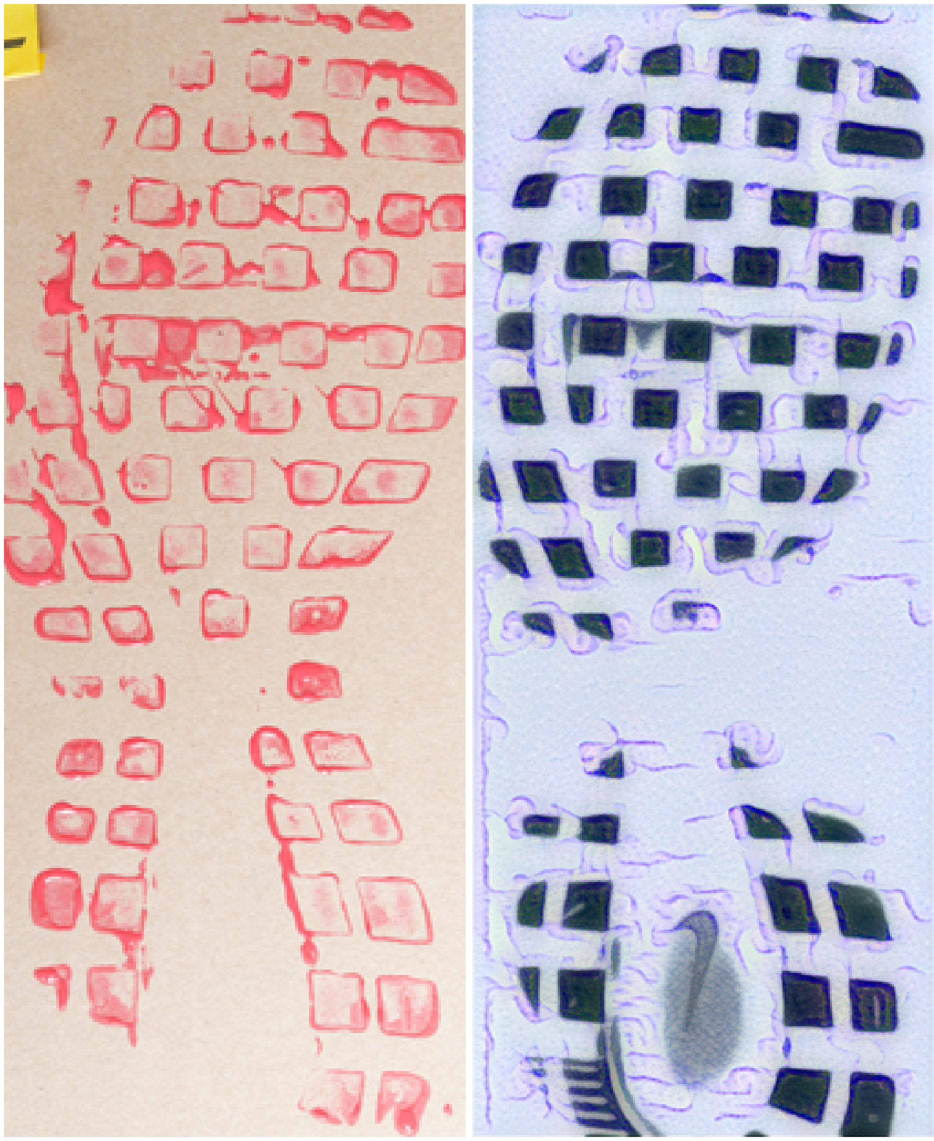
(a) Q6 02R. **(b)** Pseudo Q6.

Fig. 11. A NST of the style of Q6 to the content of 02R 01.



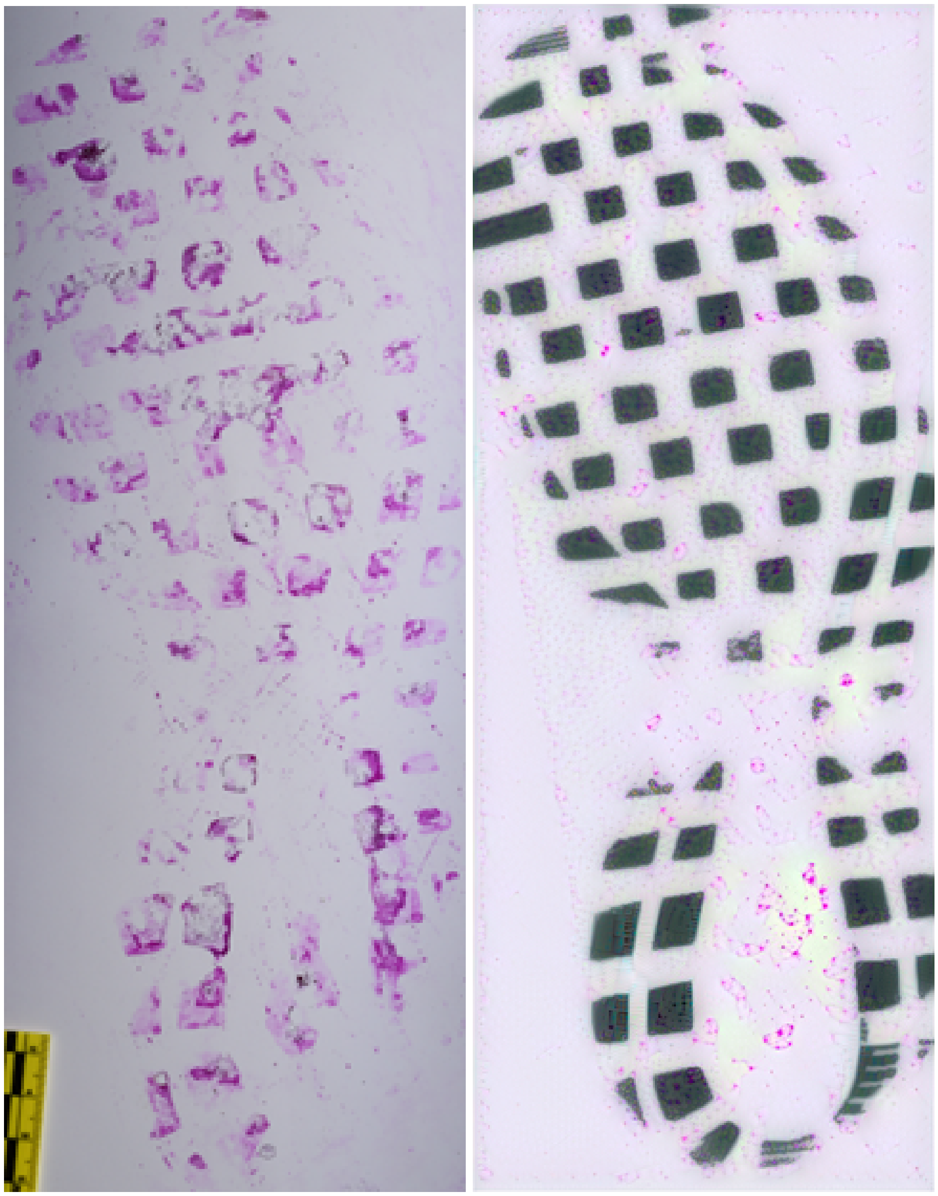
(a) Q7 03L. **(b)** Pseudo Q7.

Fig. 12. A NST of the style of Q7 to the content of 03L 01.



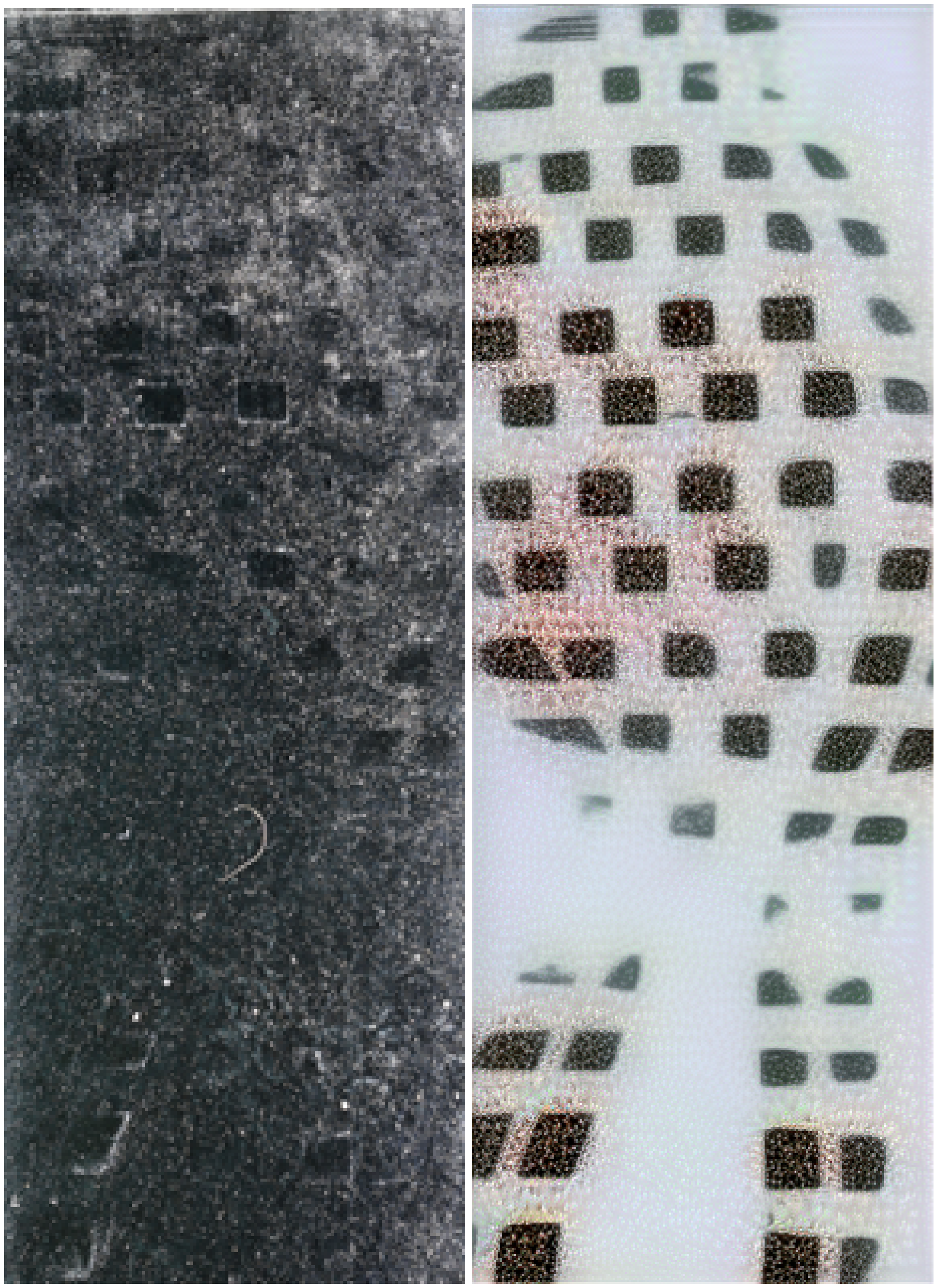
(a) Q8 03L. **(b)** Pseudo Q8.

Fig. 13. A NST of the style of Q8 to the content of 03L 01.



(a) Q9 03R. **(b)** Pseudo Q9.

Fig. 14. A NST of the style of Q9 to the content of 03R 01.



(a) Q10 03R. **(b)** Pseudo Q10.

Fig. 15. A NST of the style of Q10 to the content of 03R 01.

(a) Q11 03R. **(b)** Pseudo Q11.

Fig. 16. A NST of the style of Q11 to the content of 03R 01.

These are only observations as they appear to the eye. The neural network that uses these pseudo Qs as training images may see something else. A quantitative measure of success of the style transfer will have to be determined by measuring how well the pseudo Qs train the neural network. A practical limitation to using NST pseudo Qs as training images is that on a GPU it took about 6 minutes to complete a NST. This becomes computationally expensive when thousands of training images are desired.

4. Transferring Sharpness with NST

Here we explore our second question. Does adding a NST pre-processing step improve the comparison of a Q with a suspected K? In other words, Is it easier to distinguish the shoe that created Q from the non-matches when we compare Q to the group of Ks or when we compare an NST image of Q to the group of Ks? Neural Style transfer was applied (see Fig. 17) to each Q image and one of its associated aggregate images. For example the content of the Q1 image is combined with the style of its associated flex aligned aggregate image. Another NST combines the Q1 with the binary image, and another combines the Q1 image with its associated signed edge distance image. Neural Style transfer requires that we assign weights to the content and style images. So each of the 3 NSTs above are performed with four different weight settings: $(a,b) \in \{(.925,.1),(.425,.6),(.125,.9),(.025,1.0)\}$. We then have a total of 12 neural style transfer images for each Q from Fig. 4. Each of these will be further processed and then compared to the 30 test impressions to measure their similarity. We will also compare the original Q images to the 30 test impressions and observe the relative performance of the NST comparisons to the Q comparisons.

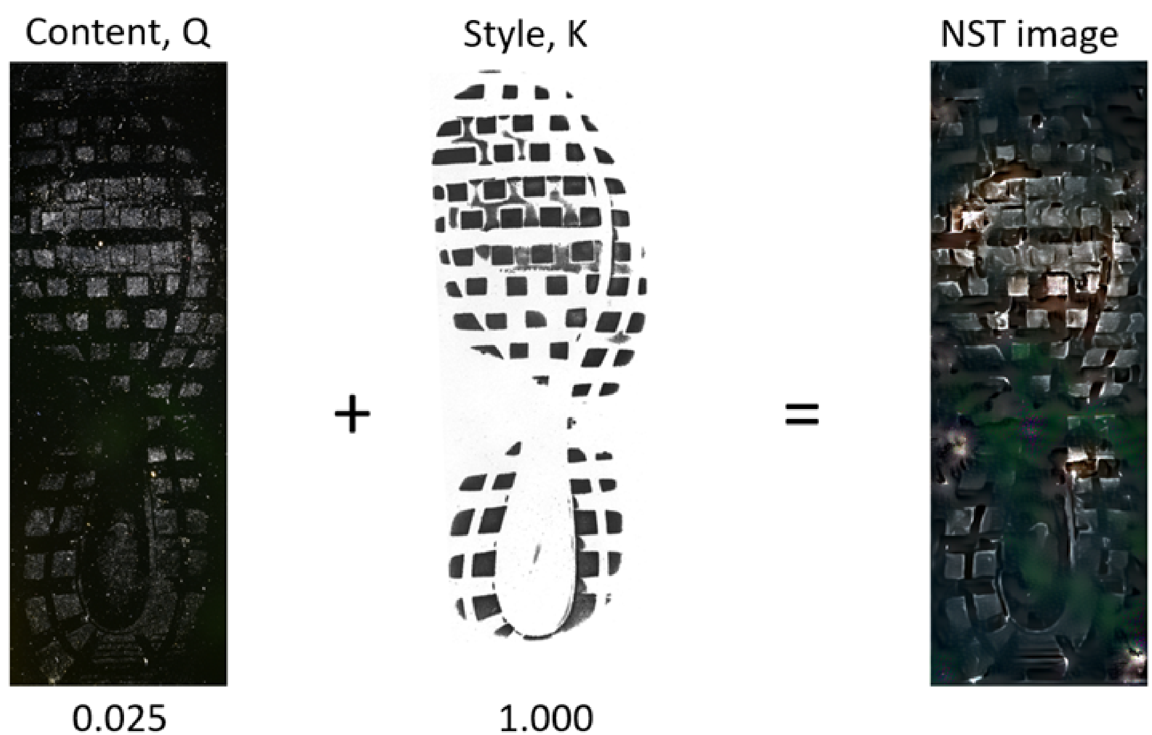


Fig. 17. A demonstration of NST on a Q and K

In Fig. 18 it is possible to see how the sharpness of the K image has been transferred to the Q image making the tread stand out. The $(.025, 1.0)$ weighting was chosen for display because this is when the effect is visually most noticeable.

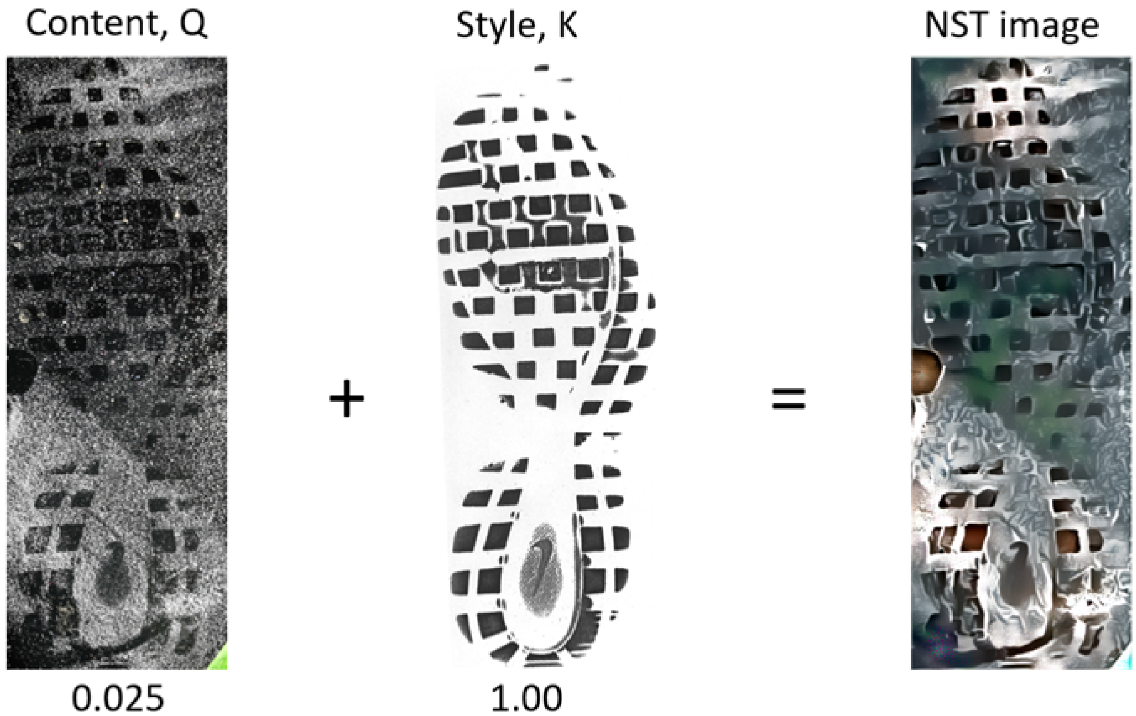


Fig. 18. A demonstration of NST on a Q and K

We perform an additional stage of pre-processing before the Q and NST images are compared to the shoe impressions. Each of the Q and NST images gets blurred and has a morphological operation [3] applied to it (see the schematic in Fig. 19). There are 4 different blur settings (corresponding to the median blur radius): 0, 5, 10, 20. The gradient settings are “no gradient”, and radius equals 2, 5 or 10. Every Q and its 12 NST images are subject to the 4x4 = 16 different settings of blurring and gradient. Each of those 16 processed images is compared to each of the 30 impressions, producing a similarity score. The maximum of the 16 scores is the one that we use. Figure 20 shows the similarity scores (r-squared) for all 13 Q1 images compared to the flex image of the 01L 01 shoe impression. The maximum score is shown in red. This is the score that will be displayed on the similarity plot, Fig. 21.

[3] We apply the medianBlur() function and morph() function from the R package, Rvision. We use k size = 0,5,10,20 for medianBlur(). We use the “gradient” option of the morph() function and the gradient size parameter settings of 2, 5 and 10. The end result of the operation is an image in which the edges of the original image are shown.

4.1. Similarity Score between Q and the Ks

For any two images, there are a variety of similarity measures [4] that can be computed to quantify the degree of correspondence between them. In this paper we perform cubic spline regression on the pixels of the Q or NST images and each of the 30 test impressions in turn. We then use the r-squared value of the cubic spline as a similarity measure between the Q image and each K, for 30 scores. The r-squared value tells us how well each pixel of the K image predicts the corresponding pixel value of the Q image. If we consider the Q image and its 30 aligned test impressions, we expect (with high probability) the 5 matching test impressions (from the shoe that created the Q image) to have higher similarity scores than the 25 impressions from the 5 close non-matching sneakers. That is, we should be able to distinguish the shoe that created the Q impression from the close non-matches. But how well does the r-squared measure allow us to discriminate between matches and close non-matches? The best it could do is have all 5 similarity scores for the 5 matches be higher than the 25 similarity scores for the 25 non-matches. The worst it could do is have all 5 scores from a non-match be higher than all 5 scores of the matching impressions. In the similarity plot in Fig. 21, the first 5 points from the left are similarity scores of the 5 shoe 01L binary test impressions with Q2. The next 5 points are from the 5 01R impressions. The 01R shoe is the shoe that created the Q2 impression. In this case the r-squared score selects the correct shoe. The next 4 groups of 5 points are scores for shoes 02L, 02R, 03L and 03R in that order. Even though the correct shoe is chosen we need to know how good the discrimination was. We need quantitative measures of discrimination. In our experiment we also compute the similarity scores of the 12 NST images with the 30 test impressions. The question we must answer is whether the NST image comparisons give us better discriminatory power than the Q image comparisons.

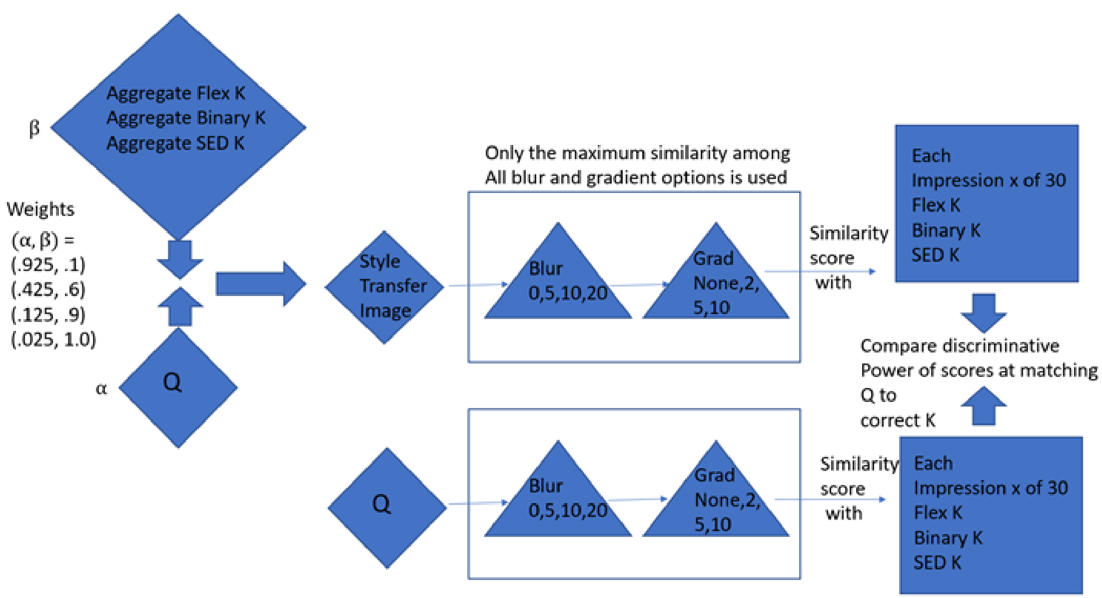


Fig. 19. A flowchart of the image comparison process.

		(0, N)	(0,2)	(0,5)	(0,10)	(5,N)	(5,2)	(5,5)	(5,10)	(10,N)	(10,2)	(10,5)	(10,10)	(20,N)	(20,2)	(20,5)	(20,10)
		[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]	[,11]	[,12]	[,13]	[,14]	[,15]	[,16]
Flex (.925, .1)	[1,]	0.325	0.250	0.229	0.163	0.448	0.110	0.130	0.141	0.466	0.161	0.185	0.192	0.395	0.212	0.236	0.233
Flex (.425, .6)	[2,]	0.288	0.172	0.157	0.118	0.381	0.040	0.047	0.055	0.403	0.061	0.074	0.079	0.334	0.083	0.095	0.093
Flex (.125, .9)	[3,]	0.282	0.164	0.152	0.116	0.373	0.032	0.038	0.048	0.396	0.048	0.060	0.066	0.326	0.070	0.082	0.080
Flex (.025, 1.0)	[4,]	0.279	0.164	0.149	0.113	0.370	0.029	0.035	0.045	0.392	0.044	0.055	0.061	0.322	0.066	0.076	0.074
Bin (.925, .1)	[5,]	0.282	0.212	0.198	0.143	0.417	0.084	0.103	0.122	0.464	0.128	0.156	0.173	0.407	0.186	0.212	0.216
Bin (.425, .6)	[6,]	0.255	0.147	0.138	0.105	0.366	0.044	0.053	0.067	0.411	0.065	0.084	0.096	0.347	0.090	0.107	0.110
Bin (.125, .9)	[7,]	0.254	0.143	0.134	0.104	0.362	0.040	0.049	0.062	0.407	0.055	0.074	0.090	0.346	0.076	0.091	0.098
Bin (.025, 1.0)	[8,]	0.253	0.141	0.132	0.104	0.361	0.039	0.047	0.061	0.406	0.051	0.070	0.086	0.344	0.072	0.086	0.093
SED (.925, .1)	[9,]	0.397	0.344	0.317	0.243	0.492	0.179	0.198	0.192	0.483	0.223	0.240	0.229	0.401	0.251	0.272	0.258
SED (.425, .6)	[10,]	0.425	0.341	0.322	0.266	0.476	0.126	0.144	0.153	0.467	0.157	0.176	0.182	0.394	0.186	0.206	0.204
SED (.125, .9)	[11,]	0.423	0.346	0.327	0.268	0.458	0.111	0.129	0.144	0.449	0.141	0.160	0.169	0.382	0.165	0.185	0.188
SED (.025, 1.0)	[12,]	0.419	0.349	0.331	0.273	0.453	0.109	0.126	0.140	0.444	0.135	0.153	0.161	0.380	0.158	0.177	0.181
Q	[13,]	0.243	0.307	0.324	0.280	0.362	0.330	0.357	0.349	0.397	0.339	0.361	0.360	0.394	0.343	0.376	0.378

Fig. 20. Similarity scores (r-squared) for all 13 Q1 images.

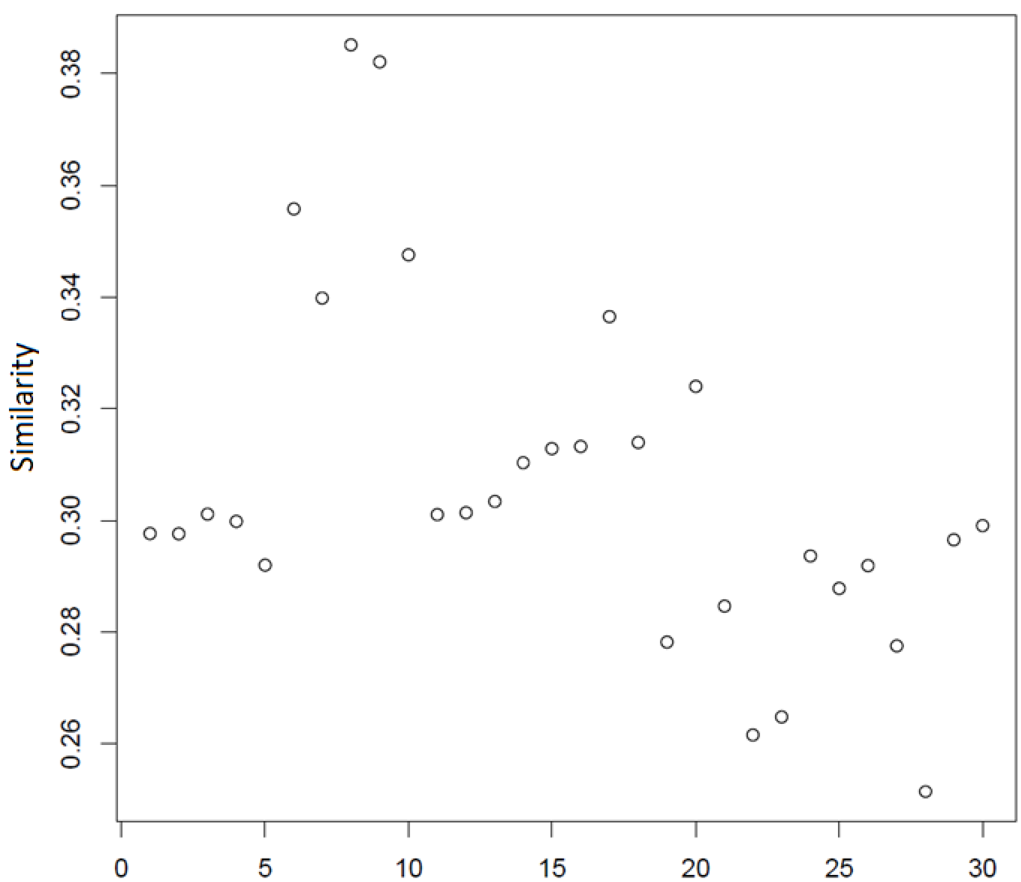


Fig. 21. Similarity plot example

4.2. Measures of Discrimination

How well do the r-squared scores discriminate between matches and non-matches? We say a matching shoe is identified correctly when its similarity score of the correct shoe is higher than the similarity scores of the non-matching shoes. The greater the separation between the match scores and the non-match scores the greater the discrimination. One way to measure this is with the sum of vertical distances of matches above the highest non-match, Fig. 22. This measure increases as the number of correct matches increases. It is also zero if there are no correct matches. Also, the sum of vertical distances increases if the separation between the matching and non-matching scores increases. In Fig. 22 there is not much difference between the lowest correct match and the highest non-match, so the vertical distance is not that great.

Another measure of discrimination is the percent correctly placed. A matching impression is correctly placed if its score is higher than the highest score of the close non-matches, and a non-matching impression is correctly placed if its score is lower than the lowest score of the matching impressions. The percent correct is 100 if the similarity scores from the 5 matching shoes are the highest. We can compare similarity plots and use the measures of discrimination to determine if the Q images or the NST images provide greater discrimination.

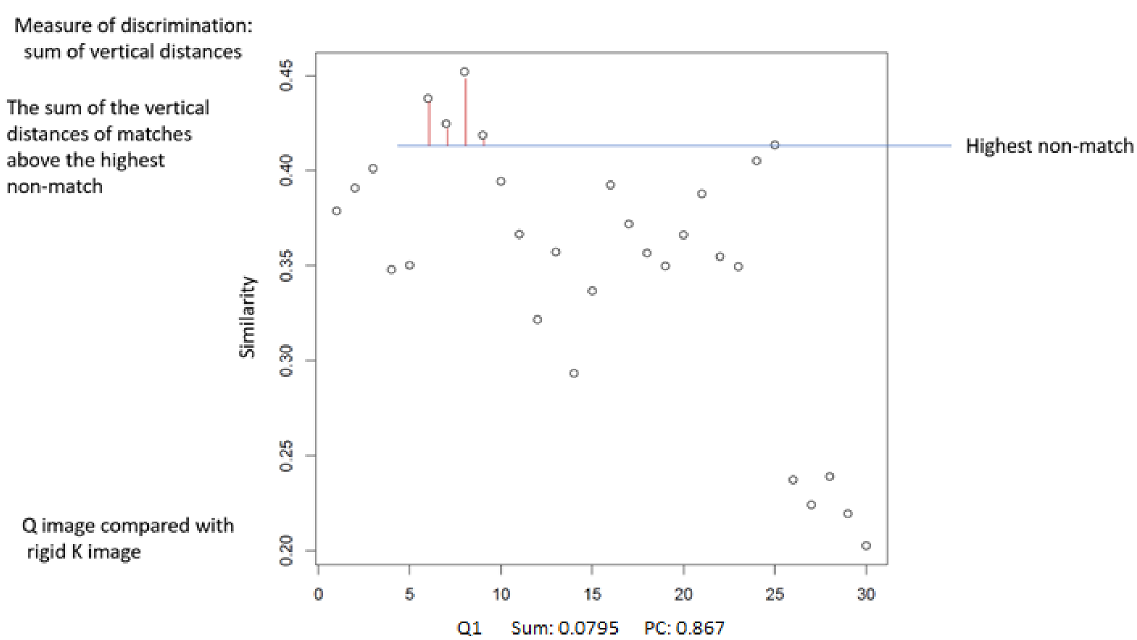


Fig. 22. Definition of sum of vertical distances.

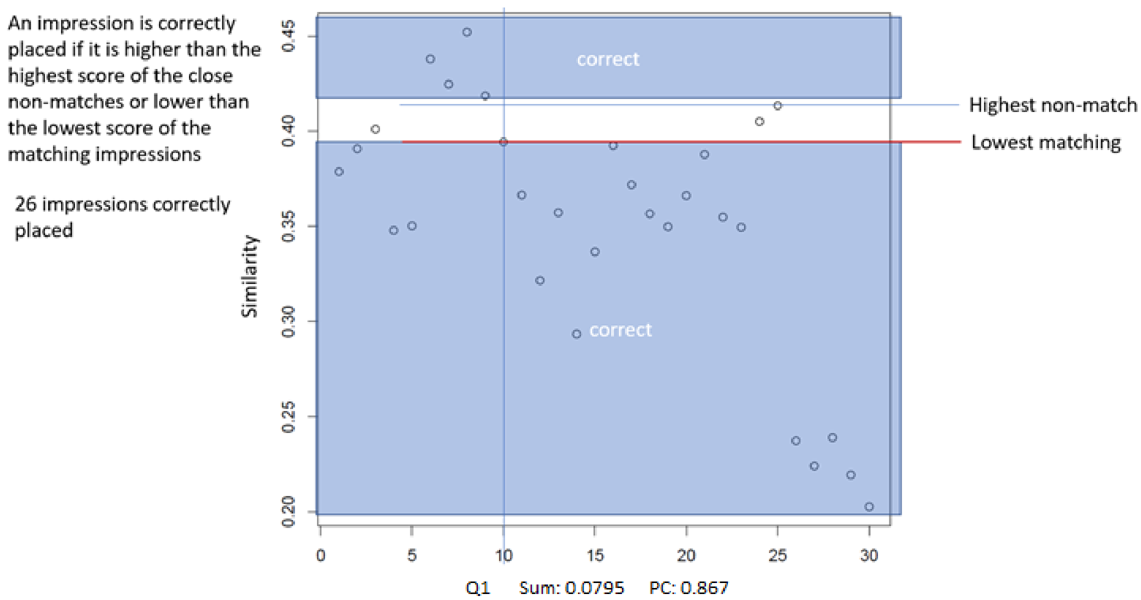


Fig. 23. Definition of sum percent correct.

4.3. Comparing Similarity Plots

In Fig. 24 the first row of plots are the similarity plots in which Q1 and the Flex Aggregate are combined via neural style transfer according to the 4 weight settings. The resulting neural style transfer image is compared via r-squared to the flex image of the 30 shoe impressions. So, for example, the first point from the left on the first similarity plot is the r-square value of the cubic spline between the NST image (with weight settings, a = 0.925, b = 0.1) of Q1 with the aggregate flex image of the 30 test impressions and the flex image of the 01L-01 image. The operation can be written as

$$compare(NST(Q1, \text{Flex Aggregate}, 0.925, 0.1), \text{Flex 01L-01}) \rightarrow R^2 \tag{2}$$

The second and third rows of similarity plots are for the Q1 image combined with the binary aggregate and SED aggregate image respectively. In the second row the NST image is compared via r-squared to the binary image of the 30 shoe impressions. In the third row the NST image is compared to the SED images. In the last row are plots of the original Q compared to the Flex, Bin and SED images of the 30 impressions.

Recall that the shoe that created the Q1 image was the 01R sneaker. The 01R shoes on the similarity plots are represented by the second group of 5 points from the left. So we hope to find that these shoes have higher similarity scores than the close non-matches. If we look at the last row, we see that when Q is compared to the flex images, 2 of the 5 01R shoes are correctly placed. The sum of vertical distances is 0.033 and the percent correct is 0.733. For the Q-Binary comparisons 3 out of 5 01R shoes are correctly placed. The sum of vertical distances increased to 0.057 and the percent correct increased to 0.833. In the Q-SED comparisons the sum of vertical distances is 0 which means that none of the matching impressions are placed correctly. The percent correct is still 0.367 because some of the non-matching impressions are placed correctly.

Now we want to determine if the NST image comparisons resulted in greater discrimination. All the NST flex comparisons have sum of vertical distances equal to 0. So none of the matches were correctly identified. The NST SED images have sum of vertical distances equal to 0 except for the (0.425,.6) weighted NST for which one shoe impression is correctly placed. Like the Q-Bin plot, the NST Binary comparisons in row 2 also show 3 shoes correctly placed, but there is slightly greater discrimination for the NST Bin (0.425,.6) comparisons which have a higher sum of vertical distances., 0.63 (compared to 0.057). So the NST comparisons are comparable to the Q comparisons. Nothing extraordinary is gained by using NST here.

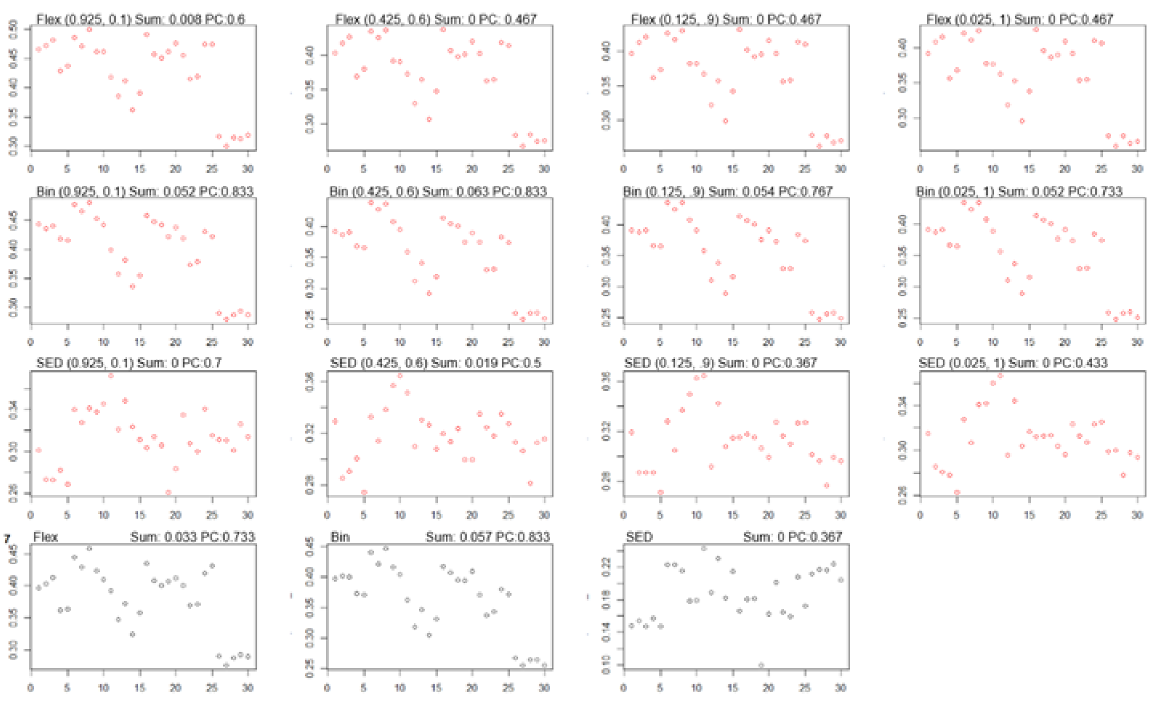


Fig. 24. Similarity plots for Q1 images

The next set of similarity plots, Fig. 25 is for the Q2 images. We first look in the bottom row at the Q images compared to the 30 K impressions. We observe that when the Q image is compared to the flex and binary versions of the K impressions there is 100 percent correct placement with sum of vertical distances 0.126 and 0.128. But the Q image compared to the SED images of the 30 K impressions does not produce good discrimination.

In fact, it looks like it is falsely identifying the 03R shoe as the matching shoe that created the Q image. We now compare the discriminatory power of the Q comparisons to the NST comparisons. We notice that for the first three Flex plots all the impressions are placed correctly. The sum of vertical distances for all three is less than the sum of vertical distances for the Q images. But the differences in the plots are small. We can say that the Q plots and the NST plots had comparable performance. We don't consider the Binary plots. We only consider the best plots which, in this case, are the Flex plots.

We want to continue comparing the Q plots to the NST plots for all the Q images. The plots for Q3, Q4, Q6, Q7 and Q8 are uninteresting. Neither the original Q images nor the NST images provide good discrimination between the matching and non-matching impressions. We leave them out in the interest of space. We display the similarity plots for the rest.

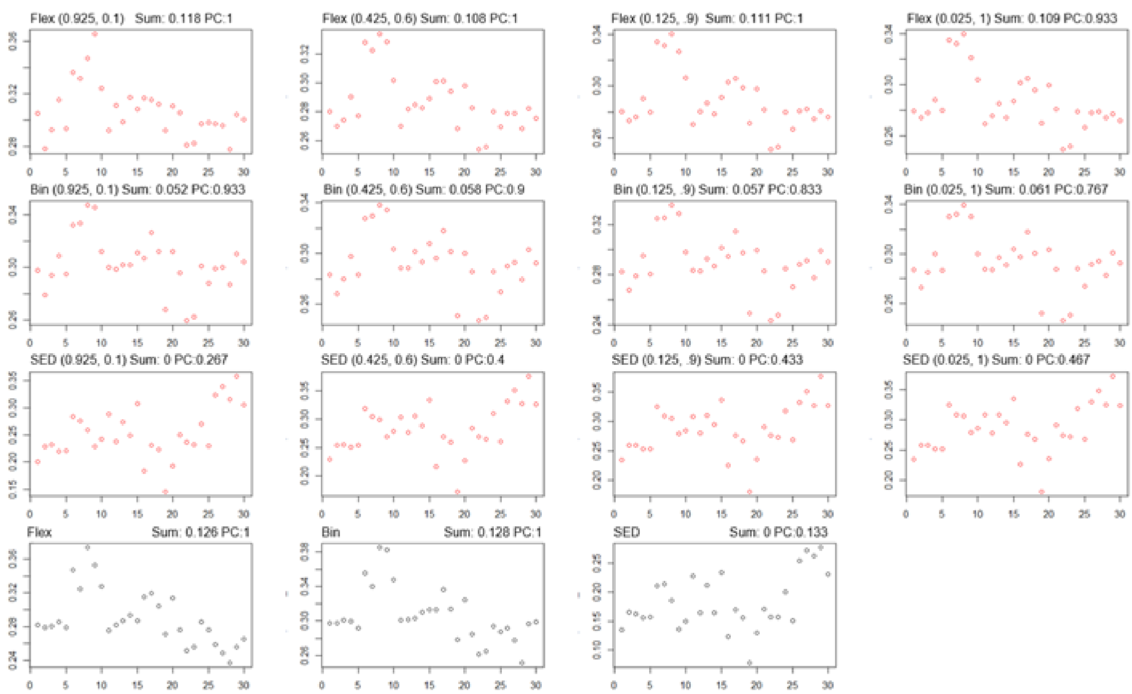


Fig. 25. Similarity plots for Q2 images

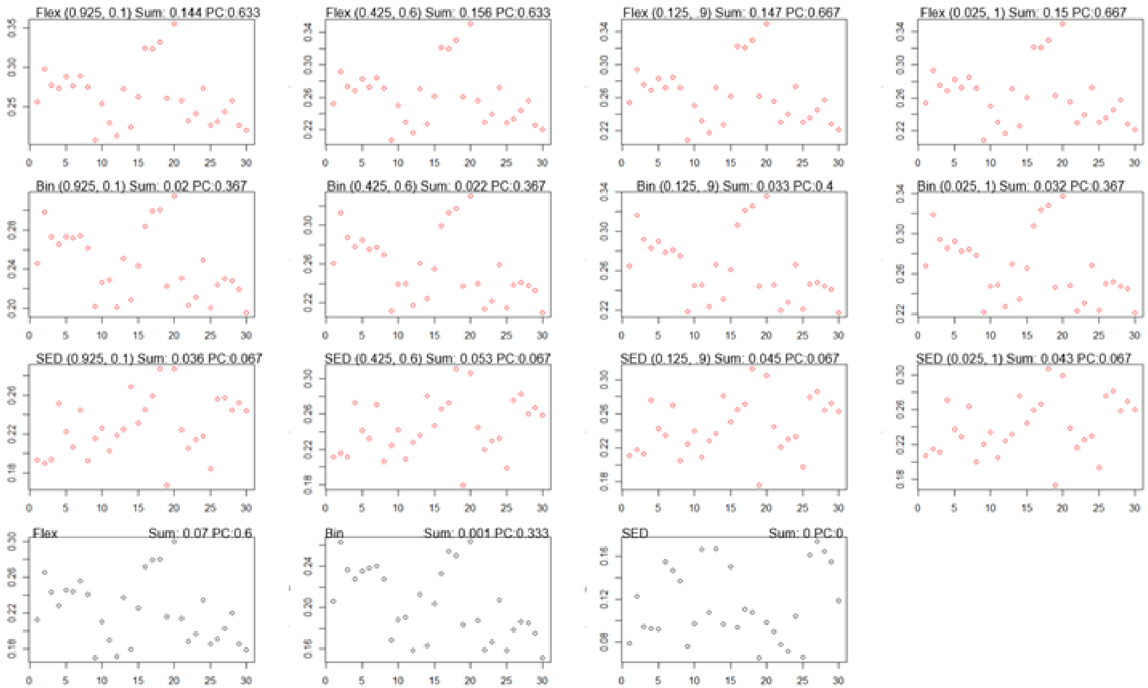


Fig. 26. Similarity plots for Q5 images

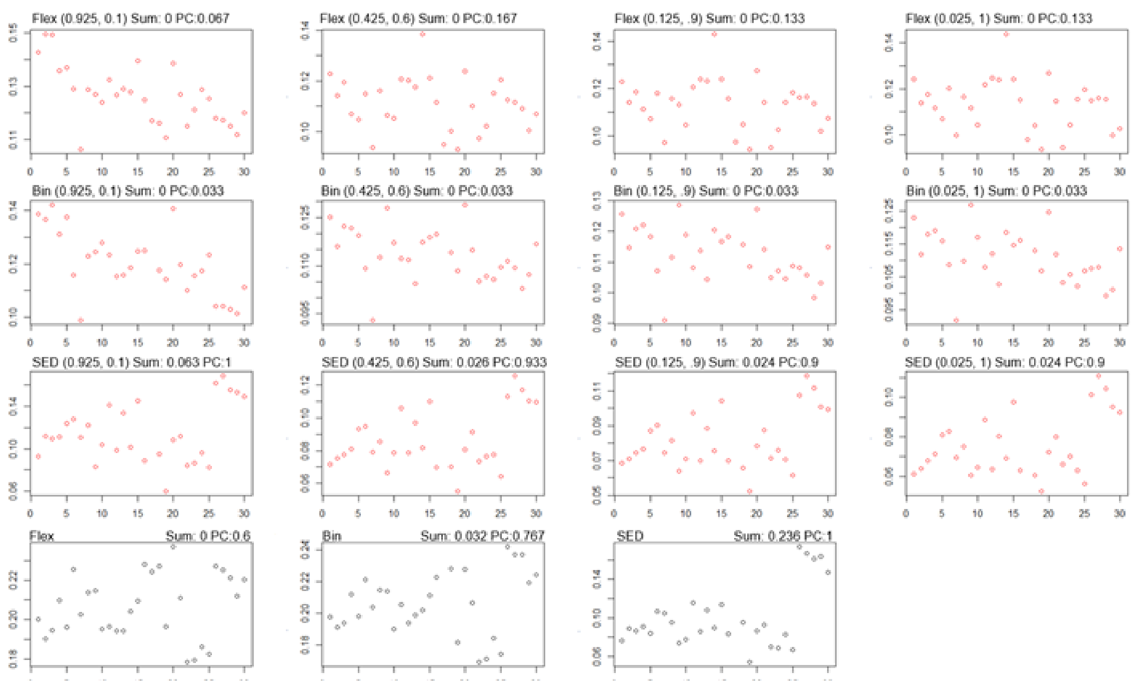


Fig. 27. Similarity plots for Q9 images

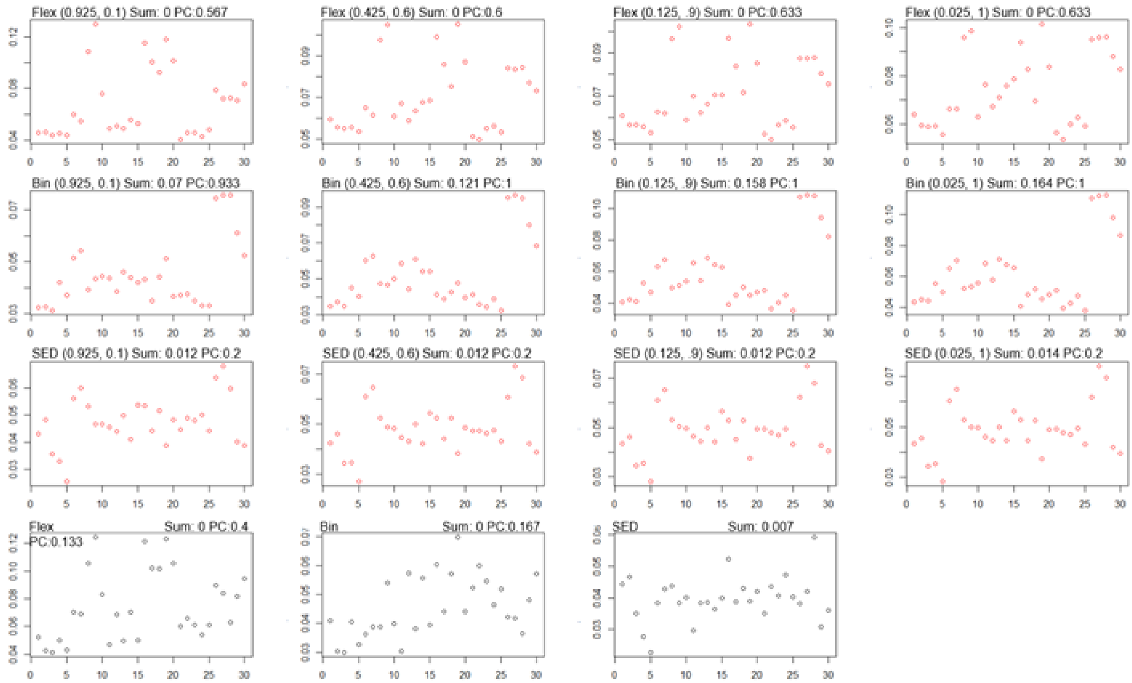


Fig. 28. Similarity plots for Q10 images

The Q5 images show that the flex NST images are better at distinguishing the matching shoe from the non-matching shoes than the Q image. This can be seen in both the sum of vertical distances and the percent correct. For the Q9 crime scene image, the original images and the NST images provide comparable discriminatory power. The most dramatic case of the NST images performing better than the original Q images is with the Q10 images. Here the original Q images provide no discrimination, but the binary NST images correctly place all 5 matching shoe impressions. For the Q11 images the performance between the NST images and the original Q images is comparable. Table 1 summarizes these results.

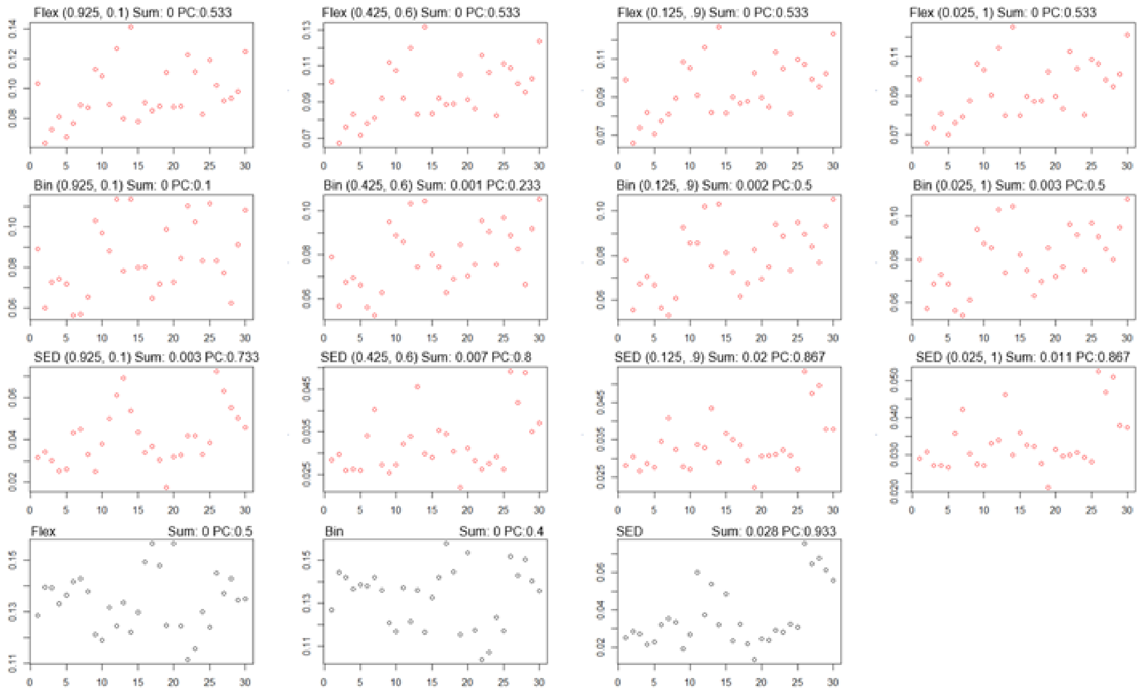


Fig. 29. Similarity plots for Q11 images

Image	Comparable	NST Superior
Q1 01R	X	
Q2 01R		X
Q5 02R		X
Q9 03R	X	
Q10 03R		X
Q11 03R	X	

Table 1. Performance of NST.

5. Conclusion

We explored the use of neural style transfer to answer two questions. In the first case we learned that NST is successful, at least visually, in transferring some of the elements of the style of a crime scene impression to a clean test impression of a shoe. A true test of success will depend on determining how well the pseudo images train a neural network. In the second case, we learned that NST can be used as a pre-processing step in comparing a crime scene image to a lineup of matching and non-matching test impressions. A crime scene image and a test impression can be combined via NST prior to comparing the crime scene image to the lineup. This step, in some cases, improves the ability to discriminate between the matching and non-matching shoes. However there are also cases where the discriminatory power is similar to comparison with the original crime scene image. We have not proposed a new process but have provided a challenge to both the Q and NST Q images. We have demonstrated that the NST version of Q had greater success in these cases.

References

[1] Chollet F, Allaire J (2018) *Deep Learning with R* (Manning Publications), 1st Ed.
[2] Venkatasubramanian G, Hegde V, Lund S, Iyer H, Herman M (2021) Quantitative evaluation of footwear evidence: Initial workflow for an end-to-end system. *Journal of Forensic Sciences* 66. https://doi.org/10.1111/1556-4029.14802
[3] Gatys LA, Ecker AS, Bethge M (2015) A neural algorithm of artistic style. *arXiv preprint* https://doi.org/10.48550/arXiv.1508.06576
[4] Venkatasubramanian G, Hegde V, Padi S, Iyer H, Herman M (2021) Comparing footwear impressions that are close non-matches using correlation-based approaches. *Journal of Forensic Sciences* 66. https://doi.org/10.1111/1556-4029.14658