

# Intra-finger Variability of Diffusion-based Latent Fingerprint Generation

## Supplementary Material

### 6. Latent Fingerprint Image Quality Metric

The scatter plot in Fig. 12 illustrates the correlation between the average Latent Fingerprint Image Quality Assessment (LFIQA) [12] scores for synthetic latent prints versus real latent prints, categorized by latent style (surface, technique). Each data point represents a specific combination of latent style from our style bank, with the bubble size indicating the volume of real style images in the dataset. Points clustered near the diagonal dashed line indicate that the generation model successfully mimicked the LFIQA quality of its real counterparts, whereas outliers reveal specific surface-technique combinations where synthetic quality significantly diverges from forensic reality.

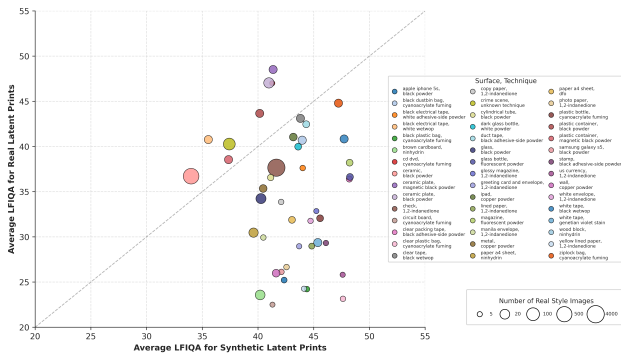


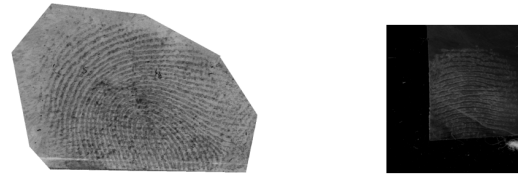
Figure 12. Scatter plot correlating the average LFIQA scores of synthetic latent against real latent prints from the same styles.

### 7. Note on Technique Categorization

Although certain entries in the summary table of the real latent datasets (Tab. 2), such as “Wetwop” and “Adhesive-side powder”, represent the same physical development technique, we categorize them separately in this study to account for variations in visual texture (see Fig. 13). Additionally, we note that not only brand formulations but also different application colors (e.g. white powder vs. black powder) may not matter since they are typically standardized to dark-on-light grayscale image by the latent fingerprint examiners. However, we chose to keep them separate to allow more examples of the same group to be synthesized.

### 8. Selection of Dataset Scale for Matching Evaluation

The decision to generate 100 synthetic exemplars per style is based on the data distribution of our aggregated style



(a) White adhesive-side powder

(b) White Wetwop

Figure 13. Visual comparison of two real latent prints developed on black electrical tape using technically identical methods. Despite the same chemical basis and surface, differences in visual appearance can be noticed.

bank. Real latent datasets vary significantly in size; for instance, the “glossy magazine + 1,2-indanedione” category contains only 20 images. We selected  $N = 100$  as a reference point to ensure statistical fairness and maximize inclusion of categories. Setting a higher threshold would force the exclusion of rare but critical style types from the one-to-one matching analysis. Consequently, Tab. 3 only includes categories that met this 100 image minimum to ensure consistency.

To specifically address the robustness and consistency of the style generation, we actually produced three distinct latent samples for each of the 100 exemplars across 45 different styles, totaling 300 images per style for the quality assessment. While space constraints in the main manuscript necessitated a condensed summary, this larger sample pool was utilized to ensure the NFIQ2 and LFIQA (Fig. 5 and 12) metrics reflect a stable and robust representation of each latent style rather than a single stochastic output.