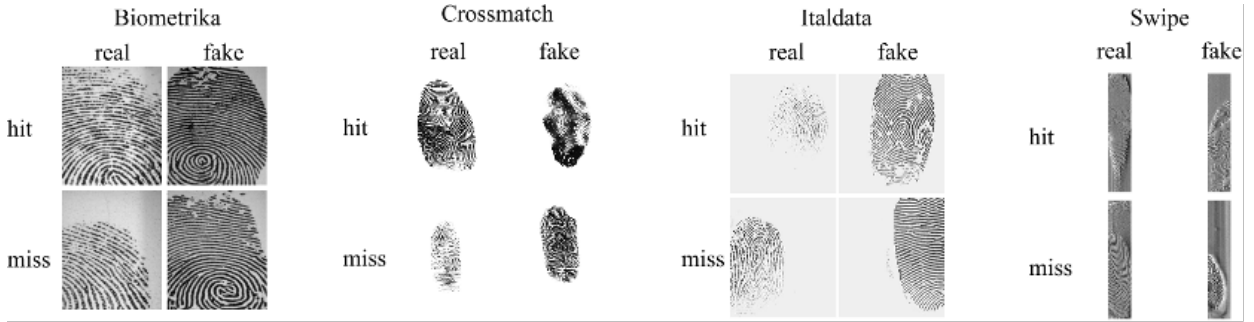


### Text-image Data Example in the Test Set

**Query:** What distinguishing features can be observed between ‘hit’ and ‘miss’ samples for real fingerprints in Fig. 5 across different benchmarks?

**Text:** Images found in these benchmarks can be observed in Fig. 5 of Section V. As we can see, variability exists not only across modalities, but also within modalities. Moreover, it is rather unclear what features might discriminate real from spoofed images, which suggests that the use of a methodology able to use data to its maximum advantage might be a promising idea to tackle such set of problems in a principled way.

**Figure:** Figure 5



**Caption:** Fig. 5. Examples of hit and missed testing samples lying closest to the real-fake decision boundary of each benchmark. A magnified visual inspection on these images may suggest some properties of the problem to which the learned representations are sensitive.

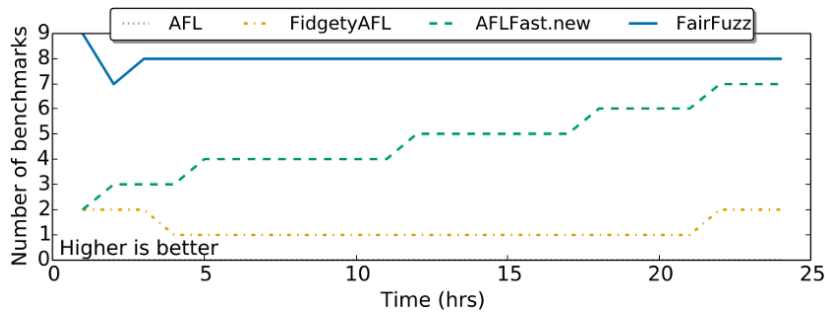
**Answer 1:** ‘Hit’ samples for real fingerprints show clearer ridge patterns and quality compared to ‘miss’ samples throughout the benchmarks.

**Answer 2:** Hit real fingerprints are clearer, more distinct; Miss real samples are blurred, less defined.

### Only-image Data Example in the Test Set

**Query:** Based on Figure 4, which fuzzing technique consistently leads in coverage across all benchmarks over the 24-hour period?

**Figure:** Figure 4



**Caption:** Figure 4: Number of benchmarks on which each technique has the lead in coverage at each hour. A benchmark is counted for multiple techniques if two techniques are tied for the lead.

**Answer 1:** FairFuzz consistently leads in coverage across all benchmarks over the 24-hour period in Figure 4.

**Answer 2:** FairFuzz, highest coverage benchmark count over 24 hours.