

ABSTRAK

Sistem pengenalan wajah banyak dimanfaatkan sebagai metode autentikasi biometrik dalam berbagai aplikasi seperti pembayaran digital, kontrol akses, dan keamanan perangkat pintar. Namun, penggunaannya diiringi ancaman serangan *spoofing*, yaitu upaya pemalsuan identitas dengan menampilkan wajah tidak asli (*spoof*) agar dikenali sebagai wajah asli (*live*). Penelitian terdahulu pada deteksi *spoofing* wajah berbasis citra *multimodal* masih didominasi pendekatan *loss Binary Cross-Entropy* (BCE) sehingga struktur representasi fitur pada ruang *embedding* belum optimal. Sementara itu, penerapan *Supervised Contrastive Learning* (SCL) yang berpotensi menghasilkan ruang *embedding* lebih terstruktur masih terbatas. Oleh karena itu, penelitian ini bertujuan untuk menerapkan SCL guna meningkatkan kualitas representasi fitur dan performa sistem deteksi *spoofing* wajah berbasis citra *multimodal*.

Metode yang diusulkan menerapkan *Supervised Contrastive Learning* (SCL) sebagai strategi utama dalam pembelajaran representasi fitur dengan memanfaatkan arsitektur *Convolutional Neural Network* (CNN) berbasis *ResNet-18* sebagai ekstraktor fitur pada dataset CASIA-SURF yang terdiri dari modalitas RGB, *infrared* (IR), dan *depth*. Representasi fitur dari setiap modalitas digabungkan menggunakan fusi *concatenation*. Pelatihan dilakukan dalam dua tahap, yaitu pembelajaran representasi menggunakan *Supervised Contrastive Loss* dan pelatihan *downstream task* menggunakan *BCEWithLogits*. Evaluasi dilakukan dengan membandingkan model SCL terhadap model *baseline* tanpa SCL menggunakan metrik APCER, BPCER, ACER, dan AUC serta dilengkapi studi ablasi.

Hasil evaluasi menunjukkan model SCL memperoleh APCER sebesar 0.28%, BPCER sebesar 0.34%, ACER sebesar 0.31%, dan AUC sebesar 0.9998, mengungguli model *baseline* dengan ACER sebesar 0.39%. Representasi fitur yang dihasilkan lebih terstruktur dengan minim tumpang tindih, didukung peningkatan *Fisher Ratio* dan *between-class distance* secara signifikan. Studi ablasi menunjukkan bahwa integrasi *multimodal*, fusi *concatenation*, serta konfigurasi parameter yang digunakan memberikan performa paling optimal. Dengan demikian, metode yang diusulkan efektif meningkatkan kualitas representasi fitur dan performa sistem deteksi *spoofing* wajah berbasis citra *multimodal*.

Kata Kunci: *face anti-spoofing*, *supervised contrastive learning*, citra *multimodal*, representasi fitur, CASIA-SURF.

ABSTRACT

Face recognition systems are widely used as a biometric authentication method in various applications such as digital payments, access control, and smart device security. However, their use is accompanied by the threat of spoofing attacks, which involve identity falsification by presenting non-authentic facial representations (spoof) to be recognized as genuine (live) faces. Previous studies on multimodal face spoofing detection are still dominated by the Binary Cross-Entropy (BCE) loss approach, resulting in suboptimal feature representation structures in the embedding space. Meanwhile, the application of Supervised Contrastive Learning (SCL), which has the potential to produce more structured embedding spaces, remains limited. Therefore, this study aims to apply SCL to improve feature representation quality and the performance of multimodal face spoofing detection systems.

The proposed method applies Supervised Contrastive Learning (SCL) as the main strategy for feature representation learning by utilizing a Convolutional Neural Network (CNN) architecture based on ResNet-18 as a feature extractor on the CASIA-SURF dataset, which consists of RGB, infrared (IR), and depth modalities. Feature representations from each modality are combined using concatenation-based fusion. The training process is conducted in two stages, namely representation learning using Supervised Contrastive Loss and downstream task training using BCEWithLogits. The evaluation is performed by comparing the SCL-based model with a baseline model without SCL using APCER, BPCER, ACER, and AUC metrics, along with an ablation study.

The evaluation results show that the SCL-based model achieves an APCER of 0.28%, BPCER of 0.34%, ACER of 0.31%, and AUC of 0.9998, outperforming the baseline model with an ACER of 0.39%. The resulting feature representations are more structured with minimal overlap, supported by significant improvements in Fisher Ratio and between-class distance. The ablation study demonstrates that multimodal integration, concatenation-based fusion, and the selected parameter configuration yield the most optimal performance. Therefore, the proposed method effectively improves feature representation quality and the performance of multimodal face spoofing detection systems.

Keywords: *face anti-spoofing, supervised contrastive learning, multimodal image, feature representation, CASIA-SURF.*