ABSTRAK

Teknologi pengenalan wajah (*face recognition*) banyak digunakan namun rentan terhadap serangan pemalsuan (*spoofing attack*), terutama *print attack* yang mudah dieksekusi. Untuk mengatasi keterbatasan sistem berbasis RGB tunggal, pendekatan multi-modalitas yang menggabungkan data RGB, *Depth*, dan Inframerah (IR) menjadi solusi efektif karena setiap modalitas menawarkan petunjuk unik untuk membedakan wajah asli dan palsu. Namun, keberhasilan pendekatan ini sangat bergantung pada strategi fusi fitur (*feature fusion*) yang digunakan, dan banyak penelitian sebelumnya membandingkan strategi ini menggunakan arsitektur yang berbeda-beda sehingga perbandingan menjadi tidak adil. Penelitian ini bertujuan untuk menganalisis secara komparatif efektivitas tiga metode fusi fitur *Concatenation*, *Addition*, dan *Cross-Attention* dalam kerangka kerja yang terkontrol. Untuk memastikan perbandingan yang adil, ketiga strategi diimplementasikan pada arsitektur *backbone ResNet-50* yang seragam dan diuji pada *dataset* multi-modal *CASIA-SURF*. Hasil penelitian menunjukkan bahwa metode *Addition Fusion* mencapai kinerja terbaik dengan *Average Classification Error Rate* (ACER) 0.68%, akurasi 99.29%, dan F1-*Score* 99.49%. Kinerja ini secara signifikan mengungguli *Cross-Attention Fusion* (ACER 0.73%) dan *Concatenation Fusion* (ACER 1.37%).

Temuan utama penelitian ini adalah superioritas metode *Addition Fusion* yang secara arsitektural lebih sederhana dibandingkan *Cross-Attention* yang lebih kompleks, yang mengindikasikan bahwa agregasi fitur komplementer secara langsung lebih *robust* dan efektif untuk tugas deteksi *print attack*. Penelitian ini menyajikan sebuah *benchmark* kuantitatif yang terkontrol dan menantang asumsi bahwa arsitektur yang lebih kompleks selalu memberikan hasil yang lebih baik.

Kata Kunci: *Face Anti-Spoofing*, Deteksi Pemalsuan Wajah, Fusi Multi-modal, Fusi Fitur, *Concatenation*, *Addition*, *Cross-Attention*, *ResNet-50*, *CASIA-SURF*.

ABSTRACT

Face recognition technology is widely used but is vulnerable to spoofing attacks , particularly print attacks, which are easy to execute. To overcome the limitations of single-modality RGB-based systems , a multi-modal approach combining RGB, Depth, and Infrared (IR) data has emerged as an effective solution , as each modality offers unique cues to distinguish between real and fake faces. However, the success of this approach is highly dependent on the feature fusion strategy employed. Many previous studies have compared these strategies using different underlying architectures, making fair comparisons difficult. This research aims to comparatively analyze the effectiveness of three feature fusion methods Concatenation, Addition, and Cross-Attention—within a controlled framework. To ensure a fair comparison, all three strategies were implemented on a uniform ResNet-50 backbone architecture and tested on the CASIA-SURF multi-modal dataset. The results show that the Addition Fusion method achieved the best performance, with an Average Classification Error Rate (ACER) of 0.68%, an accuracy of 99.29%, and an F1-Score of 99.49%. This performance significantly surpassed both Cross-Attention Fusion (ACER 0.73%) and the baseline Concatenation Fusion (ACER 1.37%).

The main finding of this study is the superiority of the architecturally simpler Addition method over the more complex Cross-Attention , indicating that the direct aggregation of complementary features is more robust and effective for the task of print attack detection. This research provides a controlled, quantitative benchmark and challenges the assumption that more complex architectures always yield better results.

Keyword: *Face Anti-Spoofing, Face Spoofing Detection, Multi-modal Fusion, Feature Fusion, Concatenation, Addition, Cross-Attention, ResNet-50, CASIA-SURF.*