ABSTRAK

Sistem pengenalan wajah saat ini sangat rentan terhadap serangan manipulasi atau *face spoofing* yang dapat mengancam integritas keamanan biometrik. Penelitian ini mengusulkan sebuah arsitektur jaringan saraf dalam untuk meningkatkan kemampuan generalisasi dalam mendeteksi *presentation attack* pada citra wajah *Red Green Blue* (RGB). Metode ini menggunakan strategi *coarse-to-fine* melalui dua cabang utama, yaitu cabang dekomposisi dan cabang fusi. Cabang dekomposisi mengekstraksi fitur *coarse* menggunakan *backbone* ResNet-18 dan fitur *fine* menggunakan *backbone* EfficientNet-B0. Mekanisme *saliency-guided suppression* diterapkan untuk menyamarkan area wajah yang dominan sehingga jaringan dipaksa untuk mengeksplorasi jejak *spoofing* yang lebih halus pada bagian wajah lainnya. Selanjutnya, fitur-fitur tersebut diintegrasikan pada cabang fusi menggunakan modul *PsiNet* yang dilengkapi dengan *Convolutional Block Attention Module* (CBAM) untuk memberikan bobot fusi secara dinamis sesuai dengan karakteristik *input*. Pengujian dilakukan menggunakan *dataset* OULU-NPU dengan pendekatan *no crop* untuk menjaga keutuhan informasi pada bingkai video. Hasil eksperimen menunjukkan bahwa model yang diusulkan mencapai performa yang sangat baik dengan nilai *Average Classification Error Rate* (ACER) sebesar 0,5% pada Protokol 1 dan 2, serta menunjukkan keunggulan signifikan pada skenario yang lebih sulit dengan nilai ACER sebesar $0{,}3 \pm 0{,}4\%$ pada Protokol 3 dan $1{,}94 \pm 1{,}41\%$ pada Protokol 4. Temuan ini membuktikan bahwa integrasi fitur *coarse* dan *fine* berbasis perhatian mampu meningkatkan ketangguhan sistem terhadap variasi sensor dan lingkungan tanpa memerlukan sensor tambahan.

Kata Kunci: *face spoofing,* CBAM, OULU-NPU*, coarse-to-fine, multi-feature fusion.*

ABSTRACT

Face recognition systems are currently highly vulnerable to manipulation attacks or face spoofing, which can threaten the integrity of biometric security. This study proposes a deep neural network architecture to improve generalization capabilities in detecting presentation attacks on Red Green Blue (RGB) face images. This method utilizes a coarse-to-fine strategy through two main branches: a decomposition branch and a fusion branch. The decomposition branch extracts coarse features using a ResNet-18 backbone and fine features using an EfficientNet-B0 backbone. A saliency-guided suppression mechanism is applied to mask dominant facial areas, forcing the network to explore finer spoofing traces in other parts of the face. Subsequently, these features are integrated in the fusion branch using a PsiNet module equipped with a Convolutional Block Attention Module (CBAM) to provide dynamic fusion weights according to input characteristics. Testing was conducted using the OULU-NPU dataset with a no-crop approach to maintain the integrity of information in video frames. Experimental results show that the proposed model achieves excellent performance with an Average Classification Error Rate (ACER) of 0.5% in Protocols 1 and 2, and demonstrates significant superiority in more challenging scenarios with ACER values of 0.3 ± 0.4% in Protocol 3 and 1.94 ± 1.41% in Protocol 4. These findings prove that the integration of attention-based coarse and fine features can enhance system robustness against sensor and environmental variations without requiring additional sensors.

Keywords: *face spoofing, CBAM, OULU-NPU, coarse-to-fine, multi-feature fusion.*