KLASIFIKASI JENIS KELAMIN BERBASIS CITRA MATA MANUSIA

DENGAN PENDEKATAN *VISION TRANSFORMER (ViT)*

Oleh

I Gde Made Hanura, 2215101062

Jurusan Teknik Informatika

Program Studi S1 Ilmu Komputer

ABSTRAK

Penelitian ini mengembangkan dan mengevaluasi model Vision Transformer (*ViT-B/16*) melalui pendekatan fine-tuning untuk klasifikasi jenis kelamin berbasis citra mata manusia. Keterbatasan wajah tertutup (masker, cadar, maupun helm) menjadi tantangan tersendiri bagi sistem pengenalan gender berbasis wajah konvensional, sehingga citra mata dipilih sebagai alternatif fitur biometrik yang tetap dapat diakses meskipun sebagian besar wajah tertutup. Dataset sekunder yang digunakan berasal dari Kaggle dengan judul "*Female and Male Eyes*", terdiri dari 11.525 citra mata (5.202 citra wanita dan 6.323 citra pria) berresolusi 60×60 piksel, dibagi dengan rasio 80:20 menjadi data latih dan data uji, kemudian data latih dibagi kembali menjadi data pelatihan akhir dan data validasi. Sebagai pelengkap, dikumpulkan pula 100 citra mata primer dari 50 responden (25 pria dan 25 wanita) melalui pengambilan data langsung. Eksperimen dilakukan dengan menguji variasi jumlah *Transformer block* yang dibuka (*unfreeze*) serta penerapan *Layer-wise Learning rate Decay (LLRD)*, dengan metrik evaluasi berupa akurasi, precision, recall, dan *F1-score*. Hasil menunjukkan konfigurasi terbaik diperoleh dengan membuka 10 Transformer block (B5) tanpa LLRD (D0), yang mencapai akurasi tertinggi sebesar 96,79% pada data sekunder dengan waktu pelatihan paling efisien (konvergen pada epoch ke-7, ±93 menit). Model yang dilatih hanya dengan data sekunder (Model A) memperoleh akurasi 96,79%, *Macro Precision* 0,9695, *Macro Recall* 0,9660, dan Macro *F1-score* 0,9675. Setelah sebagian kecil data primer ditambahkan ke dalam data latih (Model B), performa pada data sekunder meningkat menjadi akurasi 97,05%, *Macro Precision* 0,9712, *Macro Recall* 0,9693, dan Macro *F1-score* 0,9702. Namun, pengujian pada data primer menunjukkan penurunan performa yang signifikan, yaitu akurasi 61,00% pada Model A dan meningkat menjadi 62,50% pada Model B, mengindikasikan bahwa *Vision Transformer* masih menghadapi keterbatasan generalisasi terhadap variasi kondisi dunia nyata meskipun performanya sangat baik pada data dengan distribusi serupa data latih. Temuan ini mendukung pengembangan sistem klasifikasi jenis kelamin berbasis citra mata yang lebih robust terhadap kondisi data primer di masa mendatang.

Kata kunci: Citra Mata, Fine-Tuning, Jenis Kelamin, Klasifikasi, Vision Transformer

GENDER CLASSIFICATION BASED ON HUMAN EYE IMAGES USING A VISION TRANSFORMER (ViT) APPROACH

By

I Gde Made Hanura, 2215101062

Department of Informatics Engineering

S1 Computer Science Study Program

ABSTRACT

This study develops and evaluates a Vision Transformer (ViT-B/16) model through a fine-tuning approach for gender classification based on human eye images. The limitations of face-covering conditions (masks, veils, and helmets) pose a significant challenge for conventional face-based gender recognition systems, making eye images a viable alternative biometric feature that remains accessible even when most of the face is obscured. The secondary dataset used in this study was obtained from Kaggle under the title "Female and Male Eyes," consisting of 11,525 eye images (5,202 female images and 6,323 male images) at a resolution of 60×60 pixels, split using an 80:20 ratio into training and testing data, with the training data further divided into final training and validation sets. In addition, 100 primary eye images were collected from 50 respondents (25 male and 25 female) through direct data acquisition. Experiments were conducted by testing variations in the number of un*frozen* Transformer blocks and the application of Layer-wise *Learning rate* Decay (LLRD), with evaluation metrics including accuracy, precision, recall, and *F1-score*. The results show that the best configuration was achieved by unfreezing 10 Transformer blocks (B5) without LLRD (D0), reaching the highest accuracy of 96.79% on the secondary data with the most efficient training time (converging at epoch 7, in approximately 93 minutes). The model trained solely on secondary data (Model A) achieved an accuracy of 96.79%, with a *Macro Precision* of 0.9695, *Macro Recall* of 0.9660, and Macro *F1-score* of 0.9675. After incorporating a small portion of primary data into the training set (Model B), performance on the secondary data improved to an accuracy of 97.05%, with a *Macro Precision* of 0.9712, *Macro Recall* of 0.9693, and Macro *F1-score* of 0.9702. However, testing on the primary data revealed a substantial performance drop, with accuracy of 61.00% for Model A, improving slightly to 62.50% for Model B, indicating that the Vision Transformer still faces generalization limitations under real-world condition variations, despite its strong performance on data with a distribution similar to the training data. These findings support the further development of eye-image-based gender classification systems that are more robust to real-world primary data conditions.

Keywords: Classification, Eye Image, Fine-Tuning, Gender, Vision Transformer