

INTISARI

Pengembangan Model *Image captioning* untuk Membantu Aktivitas Pejalan Tunanetra di Lingkungan Perkotaan

Oleh

Syahmi Sajid
23/526178/PPA/06637

Gangguan penglihatan menjadi masalah global dengan dampak signifikan terhadap mobilitas dan keamanan individu, terutama di lingkungan perkotaan. Solusi teknologi kecerdasan buatan, seperti *image captioning*, menjanjikan bantuan bagi orang dengan gangguan penglihatan untuk membantu aktivitas sehari-hari pengidap tunanetra. Namun, ranah *image captioning* pada kasus ini masih memiliki keterbatasan dalam hal performa.

Untuk mengatasi hal ini, penelitian ini mengusulkan metode hybrid yang menggabungkan ekstraksi fitur gambar dari VGG16, ResNet50, dan YOLO pada bagian *encoder* serta LSTM dan BiGRU pada bagian *decoder* untuk menghasilkan deskripsi yang telah terbukti meningkatkan kinerja model pada Flickr8k dataset pada penelitian sebelumnya. Dengan mengadaptasi metode ini pada Visual Assistance dataset dan menambahkan proses augmentasi gambar serta *transfer learning* untuk mengatasi keterbatasan ukuran dataset, penelitian berhasil meningkatkan performa model dalam membantu aktivitas pejalan tunanetra di lingkungan perkotaan. Penelitian ini juga menguji teknik *stemming* dan penghapusan *stopwords* dalam *pre-processing* teks untuk memperkaya analisis.

Hasil evaluasi menunjukkan peningkatan signifikan dalam beberapa metrik evaluasi. Secara keseluruhan, model ini memiliki peningkatan dibanding penelitian sebelumnya. Penelitian ini mendapat peningkatan nilai BLEU-4 sebesar 60.53% dibanding penelitian sebelumnya pada Visual Assistance dataset. Keseluruhan, penelitian ini memberikan kontribusi positif dalam pengembangan solusi yang lebih efektif dan akurat bagi pengguna tunanetra di lingkungan perkotaan.

Kata kunci: *Image captioning*, Tunanetra, Lingkungan Perkotaan, Metode Hybrid, Augmentasi Gambar, *Transfer Learning*

ABSTRACT

Development of Image captioning Model for Blind Pedestrian Road Safety Navigation

By

Syahmi Sajid
23/526178/PPA/06637

Visual impairment is a global issue with significant impacts on the mobility and safety of individuals, especially in urban environments. Artificial intelligence solutions, such as image captioning, promise assistance for people with visual impairments to aid their daily activities. However, the field of image captioning in this context still has performance limitations.

To address this, this study proposes a hybrid method combining image feature extraction from VGG16, ResNet50, and YOLO on the encoder side with LSTM and BiGRU on the decoder side to generate descriptions that have proven to enhance model performance on the Flickr8k dataset in previous research. By adapting this method to the Visual Assistance dataset and adding image augmentation processes and transfer learning to address the limitations of the dataset size, the study successfully improved the model's performance in aiding the activities of visually impaired pedestrians in urban environments. This study also tests stemming and stopwords removal techniques in text pre-processing to enrich the analysis.

Evaluation results showed significant improvements in several evaluation metrics. Overall, this model has shown improvement compared to previous research. This study achieved a 60.53% increase in BLEU-4 score compared to previous research on the Visual Assistance dataset. Overall, this study provides a positive contribution to developing more effective and accurate solutions for visually impaired navigation users in urban environments.

Keywords: *Image captioning, Visually Impaired, Urban Environment, Hybrid Method, Image Augmentation, Transfer Learning*