

# Transfer Learning Model ViT-B/16 dari Dataset ImageNet ke Dataset Describable Textures Dataset dan Flowers-102 = Transfer Learning Model ViT-B/16 from ImageNet Dataset to Describable Textures Dataset and Flowers-102 Dataset

Fauzan Muhammad, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920541693&lokasi=lokal>

---

## Abstrak

Deep learning telah digunakan dalam berbagai masalah klasifikasi gambar dan deteksi objek. Dalam proses pelatihan pada Deep Learning membutuhkan data berlabel yang tidak sedikit dan akibatnya memakan waktu yang lama. Untuk mengatasi hal tersebut, metode transfer learning diperkenalkan dengan mengambil hasil pelatihan dari data sebelumnya dengan dataset yang umum, seperti ImageNet. ViT-B/16 merupakan model network yang menerapkan konsep transformer pada gambar. Model ini memiliki kinerja yang sangat baik saat dilatih menggunakan dataset ImageNet tetapi tidak terlalu baik saat dilatih menggunakan dataset Describable Textures Dataset dan Flowers-102. Penelitian ini bertujuan untuk melihat apakah metode transfer learning bisa digunakan sehingga model Vision Transformer mempunyai kinerja lebih baik saat dilatih menggunakan kedua dataset tersebut. Pada penelitian ini dilakukan transfer learning pada DTD dan dataset flowers-102 dengan menggunakan model yang sudah di pre-trained dengan ImageNet- 21K. Dari Hasil penelitian didapat bahwa model ViT mengenali dengan cukup baik pada DTD dengan akurasi pada testing hanya mencapai 76,67% dengan optimizer Adam dan dataset Flowers-102 dengan akurasi pada testing mencapai 98,91% dengan optimizer Adadelta. Model ViT ini selanjutnya dibandingkan dengan model CNN ResNet50 dan didapat bahwa hasil akurasi sedikit di bawah ResNet50.

.....Deep learning has been used in various image classification and object detection problems. The training process in Deep Learning requires quite a lot of labeled data and as a result takes a long time. To overcome this, a transfer learning method was introduced by taking training results from previous data with a common dataset, such as ImageNet. ViT-B/16 is a network model that applies the transformer concept to images. This model performs very well when trained using the ImageNet dataset but not so well when trained using the Describable Textures Dataset and Flowers-102 datasets. This research aims to see whether the transfer learning method can be used so that the Vision Transformer model has better performance when trained using both datasets. In this research, transfer learning was carried out on the DTD and Flowers-102 dataset using a model that was pre-trained with ImageNet-21K. From the research results, it was found that the ViT model recognized the DTD quite well with accuracy in testing only reaching 76.67% with the Adam optimizer and the Flowers-102 dataset with accuracy in testing reaching 98.91% with the Adadelta optimizer. This ViT model was then compared with the ResNet50 CNN model and it was found that the accuracy results were slightly below that of ResNet50.