INTISARI

Audio-Visual Convolutional Neural Networks Menggunakan *Transfer Learning* untuk Deteksi Iklan pada Video Siaran TV

Oleh

Muhammad Zha'farudin Pudya Wardana
19/448717/PPA/05800

Permasalahan deteksi iklan pada media TV memiliki tantangan yang cukup sulit karena keragamanan jenis acara dan saluran TV. Penggunaan metode *deep learning* untuk deteksi iklan telah menunjukkan hasil yang cukup baik. Namun, *deep learning* membutuhkan waktu komputasi pelatihan yang lama dengan *epoch* pelatihan yang besar untuk mendapatkan akurasi tinggi.

Penelitian ini memanfaatkan *transfer learning* untuk mengurangi waktu pelatihan dengan membatasi jumlah *epoch* sebesar 20. Fitur data video yang digunakan yaitu dari aspek audio berupa *mel-spectogram* dan aspek visual berupa *frame*. Dataset dikumpulkan dengan merekam beberapa program yang disiarkan di beberapa saluran TV nasional. *Pretrained* model CNN MobileNetV2, InceptionV3, dan DenseNet169 dilatih kembali dan digunakan untuk deteksi pada level *shot*. Kemudian dilakukan pasca pemrosesan untuk mengelompokkan *shot* pada segmen iklan dan non-iklan.

Hasil deteksi *shot* terbaik diperoleh pada model Audio Visual CNN menggunakan *transfer learning* dengan akurasi 93,26% pada 20 *epoch*, melampaui hasil CNN tanpa *transfer learning* dengan akurasi 88,17% pada 77 *epoch*. Ditambah perbaikan pada pasca pemrosesan, hasil deteksi Audio Visual CNN menggunakan *transfer learning* meningkat dengan akurasi 96,42%.

Kata kunci *– Iklan, TV, CNN, Mel Spectogram, Transfer Learning, InceptionV3, MobileNetV2, DenseNet169*

ABSTRACT

Audio-Visual Convolutional Neural Networks using Transfer Learning for Commercial Break Detection in Video Broadcasting TV

By

Muhammad Zha'farudin Pudya Wardana
19/448717/PPA/05800

The TV commercial detection problem is a hard challenge due to the variety of programs and TV channels. The usage of deep learning methods to solve this problem has shown a good result. However, it takes a long time with many training epochs to get a high accuracy.

This research uses transfer learning techniques to reduce training time and limits the number of training epochs to 20. From video data, the audio feature is extracted with mel-spectogram representation, and the visual features are picked from a video frame. The datasets were gathered by recording programs from various TV channels in Indonesia. Pretrained CNN models such as MobileNetV2, InceptionV3, and DenseNet169 are re-trained and are used to detect commercials at shot level. We do post-processing to cluster the shots into segments of commercials and non-commercials.

The best result is shown by Audio-Visual CNN using transfer learning with accuracy of 93.26% with only 20 training epochs. It is faster and better than CNN model without using transfer learning with accuracy of 88,17% and 77 training epochs. The result by adding post-processing increases the accuracy of Audio-Visual CNN using transfer learning to 96,42%.

Keywords *– Commercial, TV, CNN, Transfer Learning, Mel Spectogram, InceptionV3, MobileNetV2, DenseNet169*