

# Concept-Map Visual Encoder Actor-Critic Reinforcement Learning with Visual-Semantic Embedding-Based Reward for Image Captioning = Concept-Map Visual Encoder Actor-Critic Reinforcement Learning dengan berbasis Visual-Semantic Embedding Reward untuk Image Captioning

Lintang Adyuta Sutawika, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20522366&lokasi=lokal>

---

## Abstrak

Karya ini menggunakan encoder visual berbasis Concept-Map yang menanggulangi masalah penghubungan informasi dari citra yang telah di-encode oleh sebuah jaringan saraf tiruan konvolusional ke dalam ranah semantik yang diproses oleh jaringan saraf tiruan berbasis waktu. Pendekatan ini menggunakan komponen attention visual yang mengembangkan jaringan konvolusional sebelum dipropagasi ke jaringan berbasis waktu. Untuk meningkatkan pembelajaran cross-entropy, model dilatih dengan metode reinforcement learning dengan cara melatih value dan policy network berdasarkan jarak visual-semantic embedding distance dari representasi vector sebagai sinyal reward. Visual-semantic embedding space belajar dan menghasilkan vector space untuk citra dan teks, lalu digunakan sebagai tolak ukur kualitas suatu teks yang mendeskripsikan suatu citra. Sinyal reward membantu mengarahkan dan memaksimalkan probabilitas suatu deskripsi bagus muncul. Dataset yang digunakan adalah Flickr8k dan metric yang dilaporkan adalah BLEU-1 hingga BLEU-4

.....This work features a Concept-Map visual encoder that tackles the issue of linking encoded image information from convolutional neural networks to semantic domain processed by recurrent neural networks. The approach utilizes visual attention that extends the convolutional network before being propagated to through the recurrent network. To improve upon cross entropy learning, the model is then trained on reinforcement learning by training a value and policy network on visual-semantic embedding distance of vector representations as reward signals. The visual-semantic embedding space that jointly learns a common vector space for encoding image and caption is used to measure the quality of generated caption computing how close the vector representation is to the vector representation of the input image. The reward signal guides the policy to maximize the probability of producing good captions. The Flickr8K dataset is used and BLEU-1 to BLEU-4 for is reported.