

Captionから高解像・多フレームな動画を生成するための 深層学習モデルの改良に関する研究

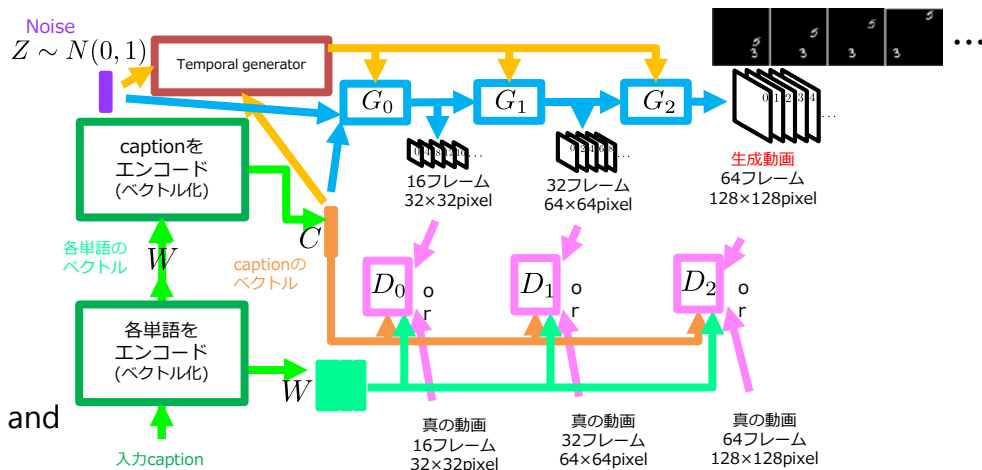
大阪大学 大学院情報科学研究科マルチメディア工学専攻 村田征矢

目的 Generative Adversarial Network (GAN)モデルを活用したCaption
(文章)からの高解像・多フレームな動画生成

内容 高解像・多フレームで鮮明な動画の生成を妨げる空間的、時間的、
意味的問題に対して、動画生成の多段階化、Temporal Generator
の導入、Caption中の単語の活用を行いGANモデルを改良

結果 “Frechet Inception Distance”と“Visual-Semantic Similarity”の指標を
用いて提案手法が既存手法 (TGANs-C) よりも、Captionに忠実で
鮮明な動画をより高解像・多フレームで生成できることを確認

利用した計算機 OCTOPUS
GPU P100 x 4基
使用メモリ 192GB



“digit 3 moves left and right, and
digit 5 moves up and down”