

進捗報告

1 DAMSM と GAN の同期学習

従来手法では DAMSM は事前学習するモデルであり、GAN モデルの訓練時には text encoder や image encoder のパラメータは固定されていましたが、先々週のゼミでの助言を受け、DAMSM と GAN を同期的に学習するようにモデルを変更してみました。

DAMSM の段階では実画像しか取り扱っていませんでしたが、GAN では生成画像も取り扱えるため、real と fake それぞれの word loss, sentence loss を DAMSM に backward するようにしています。その比率として、real での loss を 0.999, fake での loss を 0.001 かけて backward しています。

使用画像を 25 種類の鳥から 64 種類の鳥に増加させ、200epoch 程動かしました。それに伴い DAMSM の学習時のバッチサイズを 64 にしていますが、GAN の学習や画像生成時のバッチサイズはメモリの都合上変更なしです。

今現在全 200 種類すべての鳥画像をデータとして使用して動かしています。100epoch 動かす予定ですが、まだ稼働中なので 200 種での生成画像をお見せできるのは来週頃になるかと思います。

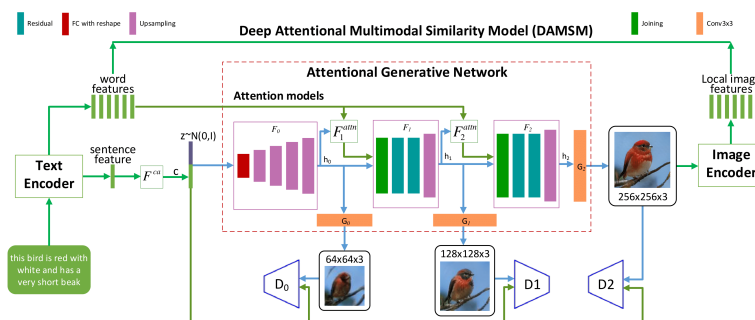


図 1: AttnGAN のモデル概要図

2 ‘概要’や‘はじめに’の下書き

創作というのは人間独自の高次な知的活動であるとされてきたが、近年の人工知能の発達が目覚ましく、その研究分野は創作の領域にまで及んでおり、計算機による創作物の理解や自動生成への試みは工学的にも興味深く大きな意義を持つようになっている。また、人工知能の発達に伴い、言語分野や画像分野などの単分野にとどまらず、多分野を複合的に取り扱うマルチモーダルな自動生成の研究も盛んになされはじめており、その一つとして AttnGAN が注目を集めている。

AttnGAN では DAMSM により単語と画像の関係を事前学習することで、テキストからの単語レベルでの画像生成を可能としている一方、単語自身の持つ自然言語的な言葉の意味合いや語法・文法などは考慮されていない。自然言語処理の手法を基に単語の持つ文化的背景や社会的文脈を考慮することで、DAMSM による Attention の精度向上や AttnGAN 全体の性能向上につながると考えられる。

そこで本研究では DAMSM に自然言語処理における Embedding 手法によって大規模コーパスを基に学習済みの単語の特徴量ベクトルを導入し、また、AttnGAN の学習時にも DAMSM を同期的に学習できるようにしたモデルを提案する。

提案手法と従来手法で得られた Attention に対して、著者が作成した比較検証用データとの類似度を SSIM に基づいて比較し、提案モデルによる Attention の精度を評価する。また、Inception Score に基づいて提案モデルによる画像生成の性能を評価する。