2021 年 5 月 28 日

B4 伊藤 薫

進捗報告

1　今週やったこと

attention map の実装

2　attention map

ViT[1] において最も識別率が高かった元画像 + グレースケール画像を訓練データとして学習した識別器の重みを使用し, attention map[2] を実装した. 明るい部分に分類するうえでの必要な画像領域に注目している. 図 1 に多義図形における元テストデータと attention map を適用した画像の例を示す.

私の主観ではあるが, 人間が多義図形だと判断している根拠と attention map が示している判断根拠は類似していることが見て取れる. このことから attention map は正しく多義図形を捉えられていることがわかり, この識別器は多義図形を識別できていると考えられる. 図 1 の最下段の attention map は主観ではあるが人間であっても多義図形だと認識しづらい画像であるにも関わらず, attention map は正しく多義図形の判断根拠を示すことができている. 今回は判断根拠を明るさで示すことで元画像と比較して画像のどの部分に注視領域があるのかを調べた. 注視領域のみを視覚化したい場合はカラーのほうがより鮮明に読み取れるのではないかと考え, カラー版 attention map は検討しておく. また, attention map を layer ごとに可視化することで過程を視覚化できるため説明する際には役立つかもしれない. 図 2 に風景画における元テストデータと attention map を適用した画像の例を示す. 風景画に関しては全体として attention がぼんやりとかかっていることが読み取れる. 一部分を注視しているわけではなく全体として風景画を認識している.

図 3 に肖像画における元テストデータと attention map を適用した画像の例を示す. 肖像画に関しては人物像の顔部分に鮮明に attention がかかっていることが読み取れる. 肖像画では主に顔に注視して判断しているとわかる.

3　今後の方針

attention map カラー版の実装, attention map を layer ごとに視覚化, 多義図形画像を探す, 別の DA の実装

参考文献

[1] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[2] VisionTransformer(ViT):VisualizeAttentionMap. `https://www.kaggle.com/piantic/vision-transformer-vit-visualize-attention-map`.

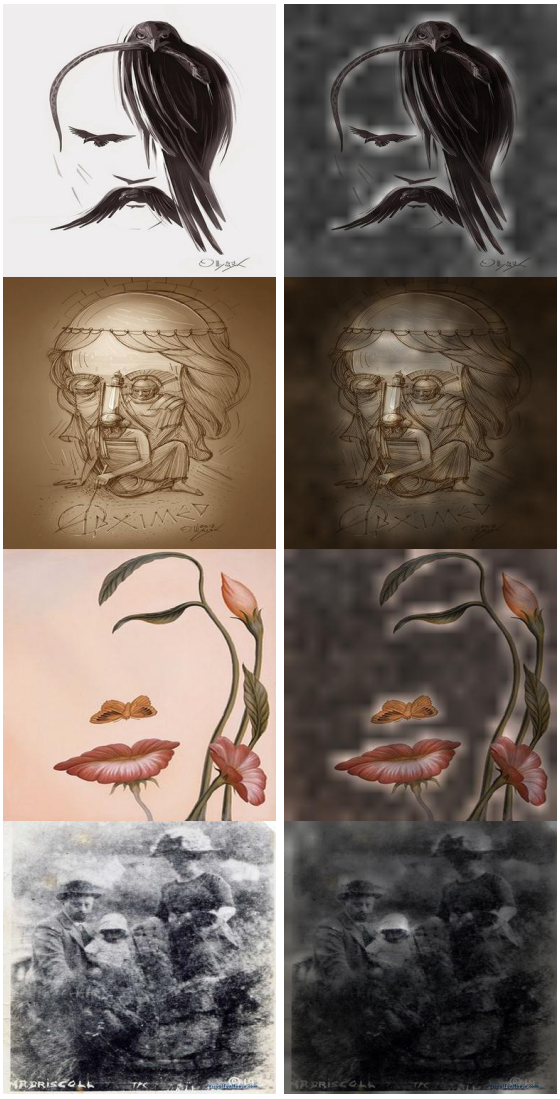

図 1: 左：元のテストデータ (多義図形), 右：attention map 適用画像

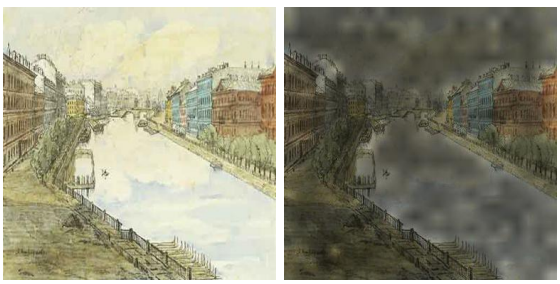

図 2: 左：元のテストデータ (風景画), 右：attention map 適用画像

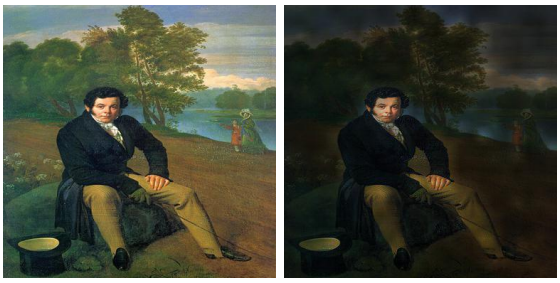

図 3: 左：元のテストデータ (肖像画), 右：attention map 適用画像