

進捗報告

1 SSIM と PSNR と RMSE による各単語にかけられた Attention の評価

前回に引き続き、SSIM と PSNR と RMSE による各単語にかけられた Attention の評価に取り組んでおり、手法による Attention の付き方を考慮させるため、輝度と範囲にバリエーションを付けて比較する方法を試している。図 1 に輝度と範囲にバリエーションを付けて作成してみた比較検証用データの一部を示す。

この比較検証用データを用いて、評価した結果を「attn_whole_relax.xlsx」と「attn_words_relax.xlsx」にそれぞれ示す。

前回までと同様に PSNR と RMSE の値も算出させているが、範囲にバリエーションを付けており、特に PSNR に関しては本来の用途と異なってくるため、Attention の評価指標としては SSIM の値を軸にしたいと考えている。

比較検証用データの輝度と範囲にバリエーションを付け、評価指標での Attention のかかり方の制約を緩めたことによって、著者による定性的・主観的評価や、全体にかかる Attention の定量的評価と、単語に対する評価とを近づけることができたと考えられる。

例えば、図 2 の 4 番や 14 番、15 番の鳥画像への Attention に対して、従来の評価指標では低い数値となっており、主観的・定性的評価と反するため、疑問に感じていたのだが、今回の手法では高い評価を与えることができた。

また、図 2 の 1 番の画像には、「this bird is brown with a white tail and has a long pointy beak」というキャプションが与えられており、「brown」と「bird」の位置はほぼ一致している。このような場合、Word2Vec や GloVe では、「brown」に対する Attention を他手法より高く付与できるのに対して、「bird」への Attention の評価は著しく低いことがわかる。その一方で、fastText や ConceptNet では、ストップワード以外の Attention がついてほしい単語に対しては、全体的にほどほどに Attention がかけられていることがわかる。

これらの一部には、昨年度の後期研究発表会でも、定性的に見られた特徴として述べたものもあるが、それらを定量的に数値として示すことができたと考えられる。

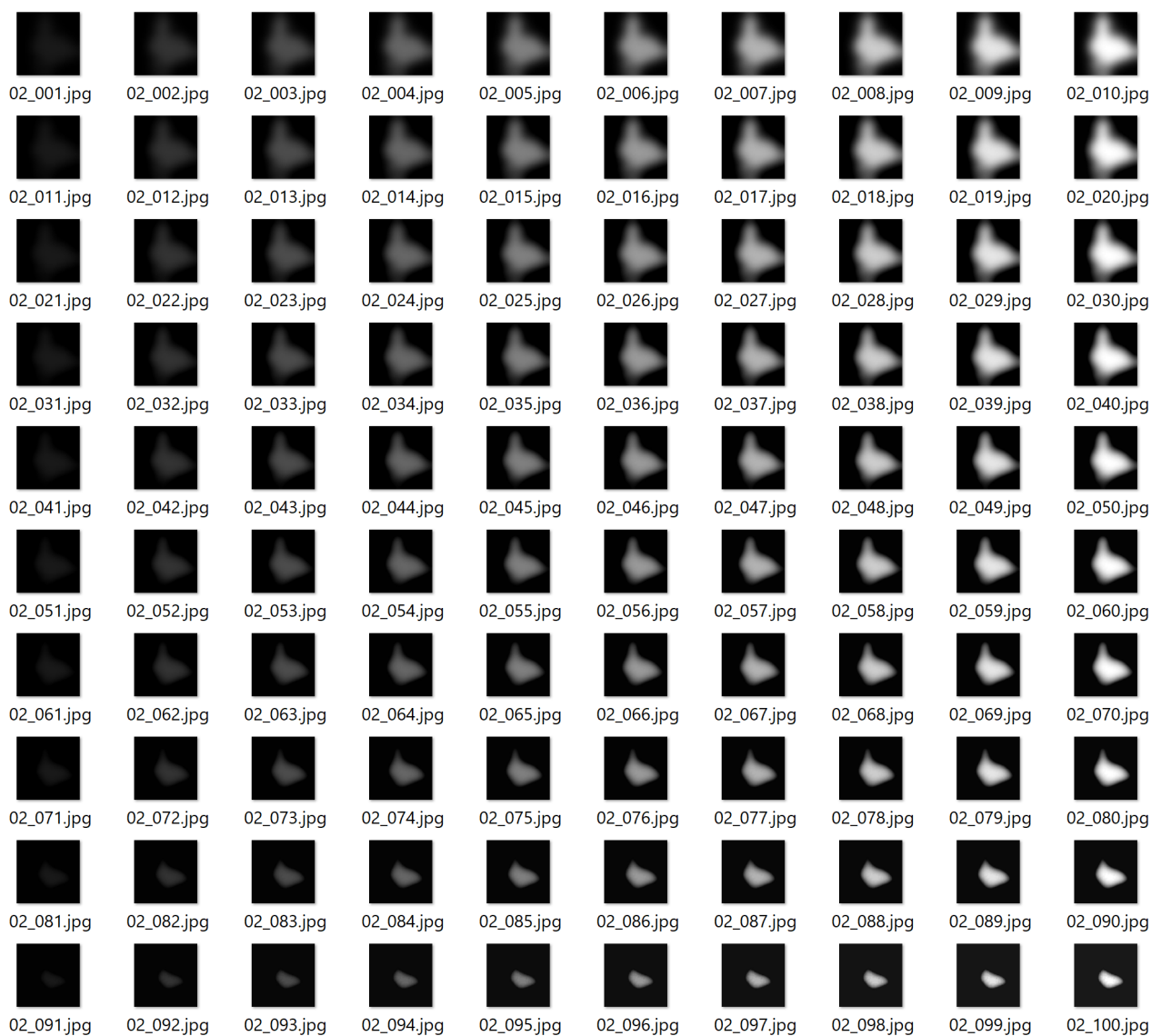
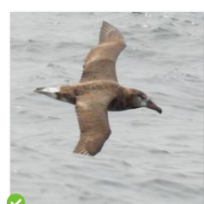
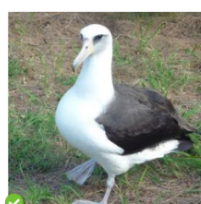


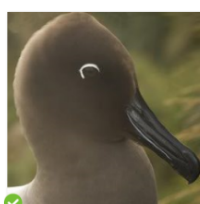
図 1: 輝度と範囲にバリエーションを付けた比較検証用データ



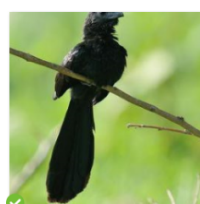
01.jpg



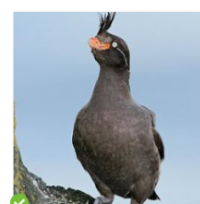
02.jpg



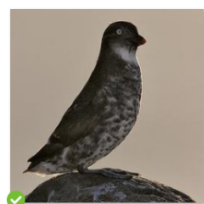
03.jpg



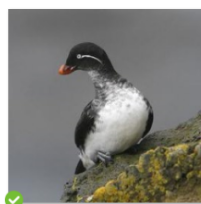
04.jpg



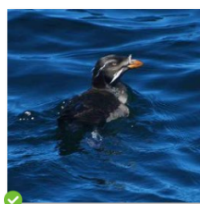
05.jpg



06.jpg



07.jpg



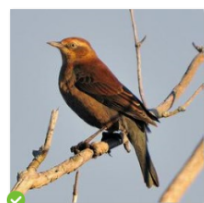
08.jpg



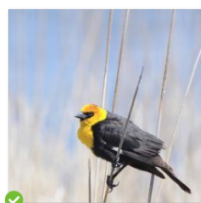
09.jpg



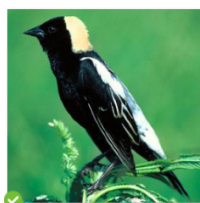
10.jpg



11.jpg



12.jpg



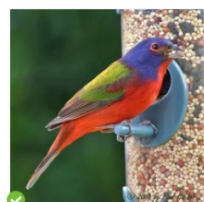
13.jpg



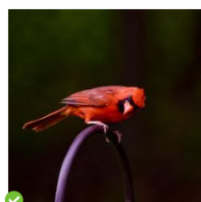
14.jpg



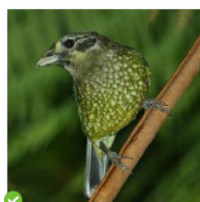
15.jpg



16.jpg



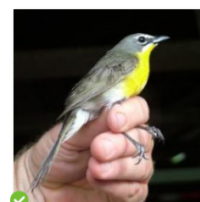
17.jpg



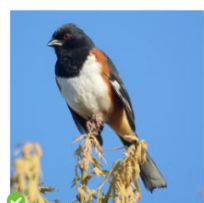
18.jpg



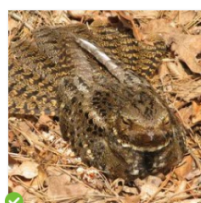
19.jpg



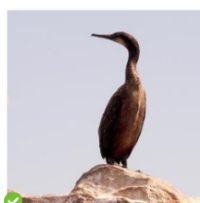
20.jpg



21.jpg



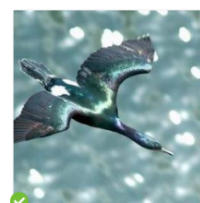
22.jpg



23.jpg



24.jpg



25.jpg

図 2: 鳥画像データ