H-039

Encoder Decoder モデルによって生成される高周波成分の有効性に関する検討

Investigation of the Effectiveness of the High-frequency Component Generated by our Encoder-Decoder Model

岡本 紗季[1)]　神野 健哉[1)]
Saki Okamoto　Kenya Jin'no

概要

我々は生成 AI や顔認識などの分野で広く利用されている畳み込みニューラルネットワーク（CNN）に着目し，CNN を用いた Encoder Decoder モデルにより，老化変換を学習させた．その結果，高齢者の特徴であるシワなどの微細な情報が失われていることを確認した．画像の品質を向上させるためには，高周波成分として知られる微細な情報が不可欠である．しかしながら，Auto Encoder のような画像サイズを小さくしながら学習をする CNN では層が深くなるほどこれらの高周波成分が失われやすい傾向がある．そこで本稿では，情報量の少ない単純な画像と CNN を含む Encoder Decoder モデルを用いて高周波成分を生成するための学習を行い，高周波成分の生成部分に関する検討を進める．

1　まえがき

年齢別に顔画像を生成する手法として，1 枚の画像から 0 歳から 70 歳まで顔の特徴を変化させずに頭部の形状の変形を可能にした顔画像が予測可能なモデルが提案されている [1]．このモデルは敵対的生成ネットワーク（GAN）を用いており，パラメータ数が多く複雑な構造を持つ．そのため簡略化した手法で年齢別の顔画像が予測できないかと考え，我々は生成 AI や顔認識などの分野で広く利用されている畳み込みニューラルネットワーク（CNN）[2] に着目した．CNN を含む Encoder Decoder モデルでも異なる年齢の顔画像が予測可能か確認するため，提案されている年齢別顔画像生成モデル [1] を用いて若者と高齢者の顔画像を生成し，データセットを作成した．図 1(a) の若者の顔画像を入力すると図 1(b) の高齢者の顔画像が出力されるように Encoder Decoder モデルを学習させた．

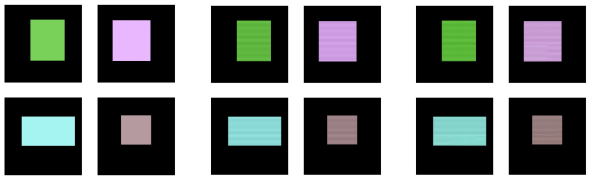
その結果，高齢者の顔画像に近づくが高齢者の特徴であるシワなどの微細な情報が失われていることを確認した．このように画像の品質を向上させるためには，高周波成分として知られる微細な情報が不可欠である．しかしながら，Auto Encoder[3] のような画像サイズを小さくしながら学習をする CNN では層が深くなるほど低周波成分の方が残りやすく，高周波成分が失われやすい傾向がある [4]．そこで本稿では，情報量の少ない単純な画像と CNN を含む Encoder Decoder モデルを用いて高周波成分を生成するための学習を行い，高周波成分の生成部分に関する検討を進める．

(a) 入力画像　(b) 正解画像　(c) 出力画像

図 1　老化変換

2　データセット

情報量の少ない単純な画像として図 2(a) に示す無地の図形と図 2(b) に示すボーダー付きの図形を用いる．背景を統一し，図形は四角形とし，ボーダーは 2 ピクセルごとに線を引いた画像を用いる．無地の図形の画像を入力するとボーダー付きの図形の画像が出力されるようモデルを学習する．モデルには Encoder で抽出した特徴を Decoder 側に結合させる Concatenation を含む．



(a) 入力画像　(b) 正解画像　(c) 出力画像

図 2　データセットの一例

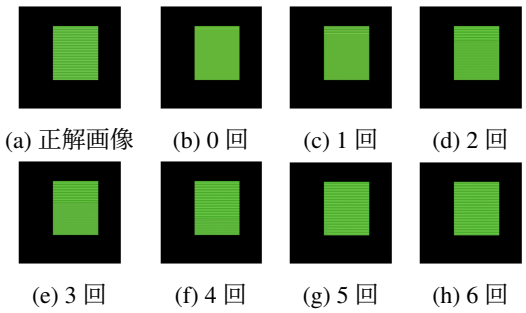
学習後に図 2(a) の画像で予測をさせた結果である図 2(c) を見ると，正解画像に近い画像が出力されていることが確認できる．

3　縮小回数が高周波成分に与える影響

Encoder と Decoder の層数を 7 層に固定し，縮小回数を変化させた実験を行い，それぞれの回数で生成される高周波成分の影響を評価する．入力画像は $256 \times 256 \times 3$ のため，例えば縦横を 1/2 にする縮小を 3 回すると 32×32 となる．正解画像と縮小回数別出力画像の例，縮小回数

1)　東京都市大学大学院総合理工学研究科情報専攻
Informatics, Graduate School of Integrative Science and Engineering, Tokyo City University

別の平均二乗誤差（MSE）をそれぞれ図 3，表 1 に示す．図 3(b) から図 3(h) のキャプションは縮小回数を表す．



(a) 正解画像　(b) 0 回　(c) 1 回　(d) 2 回

(e) 3 回　(f) 4 回　(g) 5 回　(h) 6 回

図 3　正解画像と縮小回数別出力画像の例

表 1　縮小回数別 MSE

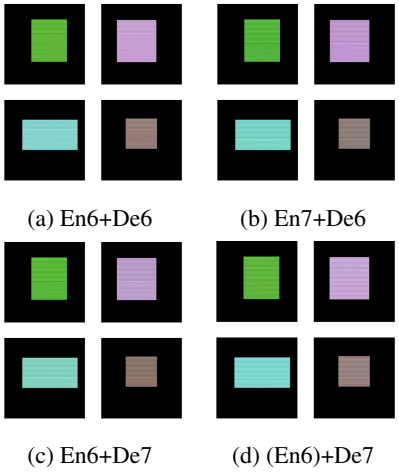
	0 回	1 回	2 回	3 回	4 回	5 回	6 回
学習	0.0039	0.0036	0.0030	0.0026	0.0021	**0.0014**	0.0015
テスト	0.0031	0.0028	0.0025	0.0020	0.0014	**0.0012**	0.0017

図 3 より，縮小回数が増えるほどボーダーの予測範囲が広がることが確認できる．また表 1 の結果からも同様のことが言え，特に 5 回縮小した場合の MSE が最も小さい．これらよりある程度は画像を縮小した方が高周波成分は生成しやすいのではないかと推測される．

4　Encoder と Decoder の役割

高周波成分の生成に対する Encoder と Decoder の役割を明らかにするため，Encoder と Decoder の層数を変更して実験を行う．前節より画像サイズを 8×8 まで縮小しながら学習をする方が MSE が小さいことから，縮小サイズを 8×8 に固定する．Encoder と Decoder の層数が 6 層の場合，Encoder が 7 層で Decoder が 6 層の場合，Encoder が 6 層で Decoder が 7 層の場合，Encoder が 6 層の学習済みの重みを用いて 7 層の Decoder のみを学習する場合の 4 つの実験を行う．それぞれの予測結果を図 4 に示す．各キャプションは学習方法を表し，「En」は Encoder，「De」は Decoder，数字は層数，「(　)」は学習済みの重みを利用していることを表す．

図 4(a) と図 3(g) を比較すると，同じ縮小サイズでも層数を重ねた方がボーダーがはっきりすることが確認できる．図 4(a) から図 4(c) の結果を見ると，図 4(a) より図 4(b) と図 4(c) の方がボーダーが明瞭に生成されていることから，層数が増えることによってより詳細な情報を抽出または再構成することが容易になったと考えられる．図 4(c) と図 4(d) より，同じモデルの構造でも Decoder のみ学習した方が図形全体にボーダーのある画像が予測可能である．これらより，高周波成分を生成するには Decoder が重要であると考えられ，Encoder と Decoder を同時に学習すると Encoder の学習状況が Decoder に影響



(a) En6+De6　(b) En7+De6

(c) En6+De7　(d) (En6)+De7

図 4　層数別予測結果

を与えていることが示唆される．また，Encoder である程度入力画像の情報が抽出でき，Decoder で画像を再構成するときに入力画像と出力画像の異なる部分に焦点を当てて学習することができた際に正解画像に近い画像が予測できるのではないかと示唆される．

5　まとめ

CNN を含む Encoder Decoder モデルを用いたボーダーを生成するシステムにおいて，高周波成分がどこで生成されているのかについて実験的に考察を行った．その結果，高周波成分を生成するためには Decoder の学習が重要であり，ある程度解像度が小さくなるまで縮小する必要があることを確認した．Concatenation の役割や Encoder がどの程度特徴を抽出できれば良いかについては，さらなる研究が必要である．今後は Concatenation の本数を変化させるなど，正解画像に近づくための Encoder の条件について検討を進める予定である．

謝辞

本研究の一部は JSPS 科研費 23K11266, 23H03387, 24K15115，東北大学電気通信研究所共同プロジェクト研究，東京都市大学重点推進研究未来知能ユニットの助成によるものです．

参考文献

[1] Roy Or-El, Soumyadip Sengupta, Ohad Fried, Eli Shechtman, “ Lifespan Age Transformation Synthesis,” Ira Kemelmacher-Shlizerman, In Proceedings of the European Conference on Computer Vision, 2020.

[2] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, ”Gradient-based learning applied to document recognition,” Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.

[3] G. E. Hinton; R. R. Salakhutdinov, “Reducing the Dimensionality of Data with Neural Networks,” Science 313 (5786), pp. 504-507, 2006. DOI: 10.1126/science.1127647.

[4] Sora Togawa, Kenya Jin’no,” Examination of the Relationship between Feature Extraction by Kernels and CNN Performance,” IEEE 2024 International Symposium on Circuits and Systems (ISCAS 2024), Resorts World Convention Centre, Singapore, May 19-22 2024.