

AutoEncoder を用いた Active Appearance Models の性能向上

渡辺 拓也† 桂田 浩一† 新田 恒雄‡ 入部 百合絵††

豊橋技術科学大学† 早稲田大学‡ 愛知県立大学††

1. はじめに

本研究では AutoEncoder を用いて Active Appearance Models (以下 AAM) [1]の精度を向上させることを目指す。AAM は主として顔画像を合成する際に用いられるモデルであり、入力パラメータを与えることで任意の顔画像を合成できる。逆に、ある顔画像が与えられた時、それに近似した画像を合成することで、その顔画像を表す入力パラメータが得られる。この操作を *fitting* と呼ぶ。Fitting 操作により得られた入力パラメータは顔画像処理において特徴点として用いられる。

AAM を構成する際、次元圧縮に主成分分析が用いられるが、主成分分析は線形空間で行われるため、非線形に分布するデータには対応できない。そこで本研究では AutoEncoder[2]を用いることにより非線形空間で次元圧縮を行い、*fitting* で得られる合成画像の精度の向上を目指す。

2. Active Appearance Models (AAM)

2.1. AAM の概要

AAM を作成する際には、図 1 のように顔画像に付与された座標と輝度の集合に対してそれぞれ主成分分析を行って形状モデル、輝度モデルを構築し、更にこれらの主成分パラメータを統合して改めて主成分分析を行うことで統合モデルを構築する。この統合モデルが AAM であり、統合モデルの主成分パラメータが AAM への入力パラメータとなる。入力パラメータから形状、輝度情報を復元することで、そのパラメータに対応した顔画像の合成ができる。

2.2. Fitting 操作について

Fitting は未知顔画像に近似した顔画像を合成する操作を指す。まず入力パラメータの初期値として乱数を AAM に与える。このとき合成された顔画像と未知顔画像のピクセルごとの輝度の平均誤差を取り、それを最小にするよう入力パラメータを最急降下法等で更新する。このようにして求められた入力パラメータが未知顔画像を表す特徴量として利用される。求められた入力パラメータを用いて合成された近似顔画像のことを、*fitting* により得

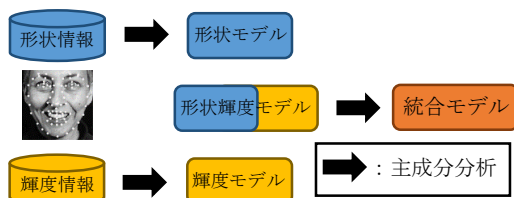


図 1 Active Appearance Models の作成

Improvement of Active Appearance Models using AutoEncoder

†Watanabe Takuya, Kouichi Katsurada, Toyohashi Univ. of Tech

‡Tsuneo Nitta, Waseda University

††Yurie Iribe, Aichi Prefectural University

られた合成画像と呼ぶ。

fitting により得られた合成顔画像は未知顔画像に近い場合が多いものの、十分な精度の合成画像が得られない場合もあった。そこで本研究では AutoEncoder を用いて AAM の性能向上を行うことで *fitting* の精度向上を目指した。

3. AutoEncoder

3.1. AutoEncoder の概要

AutoEncoder はニューラルネットワークの一種であり、図 2 のように入力層、隠れ層、出力層の 3 層からなる。入力層と出力層のノード数は等しく、また学習時の教師データとして入力層に与えられたデータと同じものを与えることで、恒等変換を学習する。このとき隠れ層のノード数を入力層より少なくすることで、隠れ層において次元圧縮が生じるため、主成分分析の代わりとして用いることが可能となる。

一般には 3 層のうち隠れ層を新たな入出力層としてその間に隠れ層を追加し、「3 層→5 層→7 層...」と多層化するが多いが、本研究では 3 層の物を中心に用いた。

3.2. ミニバッチ学習

ミニバッチ学習は複数のデータを用いて 1 回の学習を行う方法である。学習時に並列計算が可能になるため、学習の高速化に繋がるとともに過学習を抑制する効果もある。本研究ではミニバッチ学習の導入により学習の高速化と性能向上を目指す。

3.3. Pre-training と fine-tuning

AutoEncoder を始めとする深層学習では pre-training と呼ばれる教師学習を各層で予め行っておき、良い内部状態を得てから fine-tuning と呼ばれる本学習を多層のネットワーク全体に対して行うことで性能向上を図っている。本研究では 3 層の場合にも pre-training と fine-tuning を行い、性能を向上させた。Pre-training では誤差関数としてスパースな特徴抽出に優れた cross-entropy を用いた。一方、fine-tuning では一般的に誤差関数として平均二乗誤差が用いられることが多いが、correntropy と呼ばれる外れ値に強い関数を用いた MCC という誤差関数が学習に効果的であることが示されているため[3]、本研究では MCC を誤差関数として用いた。

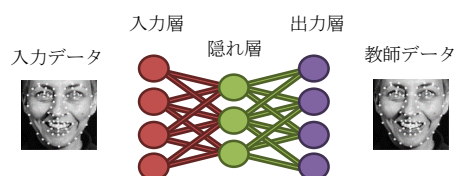


図 2 AutoEncoder の構成図

表1 各モデルの実験設定

	学習率	ミニバッチ数	学習回数
形状モデル	0.01	1	10000
輝度モデル	0.001	1	1000
統合モデル	0.01	10	10000

表2 合成画像と元画像との平均二乗誤差

AAMの構成方法	形状の平均二乗誤差	輝度の平均二乗誤差
主成分分析	6.9404	9.0698
AutoEncoder (表1の実験設定)	5.7028	7.8897
AutoEncoder (fine-tuningの導入)	5.7568	7.8557

表3 各モデルにおける恒等変換後の平均二乗誤差

	表1の実験設定	fine-tuningの導入
形状モデル	1.7523	1.6967
輝度モデル	9.6204	9.3946
統合モデル(形状)	2.6432	2.5040
統合モデル(輝度)	10.1273	10.0460

4. 評価実験

4.1. 実験データ

実験データとして研究室学生の顔画像を中心とする630枚の画像データを用いた。顔画像の解像度は640x480pixel, 8bitのグレースケール画像である。

4.2. 実験概要

実験では、まず fitting による合成画像を生成し元画像との形状、輝度の平均二乗誤差を取ることで精度の評価を行った。また、形状モデル、輝度モデル、統合モデルの AutoEncoder 導入による次元圧縮の性能についても元画像の形状、輝度との平均二乗誤差を取ることで確認した。入力次元数は形状モデル148次元、輝度モデル7159次元、統合モデル462次元である。評価は600枚を学習データ、30枚をテストデータとして交差確認法によって行った。

4.3. 結果と考察

表1に各モデルの次元圧縮の実験において最も良い結果が得られた学習設定を示す。なお、輝度モデルの実験の学習回数が少ない理由は、次元数が大きく十分に学習を進められなかったためである。表2に fitting によって得られた合成画像の精度を示す。表より、主成分分析を用いてAAMを構成した場合に比べ、表1の実験設定で学習した AutoEncoder 組み込んだ AAMの方が精度の高い画像を合成できていることがわかる。これは、AutoEncoder を用いることで線形空間だけではなく非線形空間のデータにも対応できる表現力を持ったためであると考えられる。ここで図3に fitting 操作によって得られた合成画像を示す。図3(b)を見ると、目の領域において目が潰れておりうまく合成できていないことがわかる。これに対して、図3(c)は AutoEncoder を用いることで主成分分析を用いた手法に比べ目の周辺を適切に合成できていることがわかる。これは、目のような小さい領域では値の変動による全体への寄与が小さいため、線形空間での次元圧縮では学習が適切に行われなかったためであると考えられる。AutoEncoder により非線形空間での次元圧縮を行うことでこのような小さい領域にも対応できたと考えられる。

次に fine-tuning の効果について考察する。表2より、fine-tuning を行った場合と行わなかった場合で合成画像の精度がほとんど変わっていないことが分かる。そこで、



(a)元画像 (b)主成分分析 (c)AutoEncoder

図3 fittingにより得られた合成画像

fine-tuningの導入により AutoEncoder の性能が変化しているかどうかを調べた。表3に各モデルの恒等変換の性能を示す。これは、未知顔画像を入力として AutoEncoder による恒等変換を行い、出力された画像と元画像の形状と輝度の平均二乗誤差を算出したものである。表3より fine-tuning を用いることで恒等変換の性能は向上していることがわかる。以上から、fine-tuning の導入により AAM の性能は向上しているにも関わらず、fitting により得られた合成画像の精度は向上していないことがわかった。これは fine-tuning の際に誤差関数を変更することで AutoEncoder の内部構造は変化した、fitting 操作が AutoEncoder の内部構造の変化に対応できなかったためであると考えられる。

5. まとめ

本報告では非線形な次元圧縮を行う AutoEncoder を主成分分析の代わりに用いることでAAMの性能が向上しさらに fitting による合成画像の精度が向上することを示した。しかし、fine-tuning を導入した際に fitting によって得られる合成画像の精度が向上しないという結果も得られた。これは誤差関数を変更したことによる AutoEncoder の内部構造の変化に fitting が対応できなかったためであると考えられる。

今後は AutoEncoder の内部構造を考慮した fitting の手法について検討したい。また、本研究によって AAM で AutoEncoder を用いることの有効性は確認できたが、AutoEncoder の学習には今回導入したミニバッチ学習や fine-tuning の他にも様々な学習方法が提案されている。今後はそれらの学習を行うことによる AAM の性能向上についても取り組みたい。

参考文献

- [1] T. F. Cootes, et al: Active Appearance Models, IEEE Transactions on PAMI, 23(6), pp681-685(2001)
- [2] G. E. Hinton* and R. R. Salakhutdinov Reducing the Dimensionality of Data with Neural Networks, Science, Vol.313, pp.504-507(2006).
- [3] Y. Qi, Y. Wang, and X. Zheng, "Robust Feature Learning by Stacked Autoencoder with Maximum Correntropy Criterion", ICASSP, 2014 pp.6766-6770