H-030

中間制約損失を用いた歪み付き画像認識
Distorted Image Recognition using Intermediate Constraint Loss

鈴木 聡志†‡ 谷田 隆一† 木全 英明† 庄野 逸‡
Satoshi Suzuki Ryuichi Tanida Hideaki Kimata Hayaru Shouno

1. はじめに

Deep Neural Network (DNN) は画像認識に対して高い性能を示すが，ノイズやブラーなどの画像歪みに対して脆弱である事が知られている[1]．この脆弱性に対処するために，Fine-tuning に基づく手法が複数提案されている．Fine-tuning は歪みの無い画像で学習済みの DNN を 歪み画像とそのラベルを用いて再度学習する手法である．Fine-tuning は歪み画像認識に対して有効である事が知られているが，DNN のモデルパラメータを大幅に書き換え，中間層の反応パターンを変更してしまう可能性がある．一般に，DNN の深い階層のニューロンは高次の画像特徴に反応するため，Fine-tuning の前後で反応のパターンが変更してしまう事は学習効率の点で望ましくないと考えられる．Fine-tuning の前後でニューロンの反応パターンを一貫したものにする事で，学習済みの DNN の獲得している画像特徴への反応パターンを保持する効率的な学習が可能になる事が期待される．

本研究では．歪み付き画像認識のための新しい損失関数である“中間制約損失”を提案する．この損失は，Fine-tuning 対象の DNN の深層の反応パターンを基準となる応答パターンに一致させるように制約を課す．提案手法では Fine-tuning されていない DNN に歪み無し画像を入力したときの反応パターンを基準として採用し，Fine-tuning の前後で一貫した深い階層の反応パターンを実現する．標準的な画像認識データセットである ImageNet2012 を用いた検証実験の結果，Fine-tuning に中間制約損失を導入して学習した DNN は，従来の Fine-tuning のみで学習を行った DNN よりも高い歪み画像認識精度を示す事が明らかになった．さらに，反応パターンを制約せず，基準との平均二乗誤差を用いたベースラインと比較しても高い歪み画像認識精度を示した．

2. 従来技術:Fine-tuning による歪み付き画像認識とその課題

DNN の画像歪みに対する脆弱性に対して Vasiljevic ら[2]や Zhou[3]らは Fine-tuning を用いた手法の有効性を示した．Fine-tuning は歪み画像を用いて学習済みの DNN を再学習する手法であり，先行研究では DNN に対して一定のロバスト性を与える事が示唆されている．近年もいくつかの歪み付き画像認識手法が提案されているが，多くの手法が DNN の Fine-tuning を発展した手法である[4][5].

しかし，前述の通り，Fine-tuning は単純に歪み画像とそのラベルを用いて DNN を学習するものであるため，DNN が従来の学習によって獲得した特徴抽出機構を書き換えてしまう可能性がある．一般に，学習済みの DNN の深層に存在するニューロンは高次の画像特徴に選択的に反応する事が知られている[6]．つまり，Fine-tuning はこれらのニューロンが反応する特徴を学習の過程で書き換えてしまう可能性がある．本研究では，上記の Fine-tuning の性質を鑑みて，高次の画像特徴に反応する深層のニューロンの反応パターンは Fine-tuning の前後で変更されるべきではなく，これらが一貫したものであれば効率的な学習が可能である，という仮説を基に，中間制約損失を用いた歪み付き画像認識手法を提案する．

3. 提案手法

Fine-tuning による反応パターンの変更を防ぎ効率的な学習を実現するため，中間制約損失を提案する．中間制約損失は歪み画像に対する DNN の反応パターンを基準となる反応パターンに一致させる損失である．本研究では，基準となる反応パターンとして Fine-tuning 前の DNN に歪み無画像を入力した際の反応パターンを採用した．

$h^i_{undist.}$と$h^i_{dist.}$をそれぞれ，歪み無し画像，歪み画像を入力した際のi番目の DNN の中間層の出力とする．中間制約損失の基本的なアイデアは，$\mathbf{1}_{>0}(h^i_{undist.})$と$\mathrm{step}(h^i_{dist.})$を一致させる二値分類タスクとして定義する点にある．ここで，$\mathbf{1}_{condition}(\cdot)$は条件 condition が真なら 1 を，偽なら 0 を出力する関数であり，$\mathrm{step}(\cdot)$は下記のステップ関数を示す:

$$\mathrm{step}(x) = \begin{cases} 1, \text{if } x > 0 \\ 0, \text{if } x \leq 0 \end{cases}.$$

ステップ関数は$x = 0$で微分不可能点を持ち，微分不可能点以外の領域では微分値が 0 となるため，勾配を各層で計算しつつ誤差情報を逆伝播していく DNN の学習には適さない．そこで，本研究ではステップ関数の代わりに全てのxの値域で微分可能な関数を用いる事で，勾配情報に基づく DNN の学習を実現する．具体的には，ステップ関数と同様に，$x = 0$付近でカーブを描くシグモイド関数をステップ関数の代用とする:

$$\mathrm{sigmoid}(x) = \frac{1}{1 + \exp(-\alpha x)}.$$

また，ステップ関数をよく近似するために，シグモイド関数のカーブの傾斜を決定するゲインパラメータαを10^2とした．最終的に，二値分類タスクの結果を一致させるための損失関数としてクロスエントロピーを用いて損失を算出する事で，中間制約損失$Loss_{inter}$を得る:

$$Loss_{inter} = \frac{1}{N}\sum\left(-u^i \cdot \log d^i - \left(1 - u^i\right) \cdot \log\left(1 - d^i\right)\right).$$

ここでは簡単のために，$u^i = \mathbf{1}_{>0}(h^i_{undist.})$，$d^i = \mathrm{sigmoid}(h^i_{dist.})$として書き換えた．なお，中間層の出力$h^i_{undist.}$と$h^i_{dist.}$は ReLU 等の非線形変換を施す前の値を用いる．

DNN は softmax cross-entropy 等の画像認識用の損失を示す$Loss_{recog.}$と中間制約損失$Loss_{inter}$の線形和:

$$Loss_{final} = Loss_{recog.} + \lambda \cdot Loss_{inter}$$

† NTT メディアインテリジェンス研究所
NTT Media Intelligence Laboratories
‡ 電気通信大学
University of Electro-Communications

表 1　歪み付き画像認識における，提案手法とベースラインの比較.

	Undistorted	Noise	Blur
Fine-tuning	0.6744	0.5660	0.5073
MSE	0.6713	0.5740	0.5119
Proposed	**0.6769**	**0.5769**	**0.5129**

によって学習される．ここでλは中間制約損失の強度を決定するハイパーパラメータである．通常の Fine-tuning は上記の$Loss_{final}$において$\lambda = 0$の場合と同値であり，中間層の反応パターンについて制約を課していない手法である事がわかる．

4. 実験

本研究では，自然画像認識データセットである ImageNet 2012 を用いて提案手法の有効性を検証する．画像歪みにはガウシアンノイズとガウシアンブラーを採用し，ガウシアンノイズの標準偏差σ_nを$\sigma_n \in \{20,40,60,80,100\}$，ガウシアンブラーの標準偏差を$\sigma_b$を$\sigma_b \in \{2,4,6,8,10\}$とした．また，ガウシアンブラーのカーネルサイズは$4 \times \sigma_b - 1$とした．DNN モデルには ImageNet に対して有効である事が知られている VGG-16 を用いた．深層のニューロンの反応パターンとして 14 層目，15 層目の全結合層のニューロンを$h^i_{dist.}$, $h^i_{undist.}$の算出に用いた．

提案手法と比較するためのベースラインとして中間制約損失を用いない通常の Fine-tuning と，3 節で定義した DNN の中間出力$h^i_{undist.}$，$h^i_{dist.}$に対して平均二乗誤差(MSE)損失を用いて制約を与える手法を用いる．提案手法を含めていずれの手法も最適化手法に誤差逆伝播法で算出した勾配に基き，確率的勾配降下法を用い，学習率は 0.001，バッチサイズ 64 で 40 エポック学習した．また，ハイパーパラメータλは実験的に$\lambda \in \{0.125,0.1,0.075,0.05,0.025\}$から選択し，提案手法では 0.1，MSE 損失では 0.05 を選択した．

表 1 に，提案手法，ベースラインの Fine-tuning，および MSE を用いた中間層の制約の 3 手法における，歪み付き画像認識結果を示す．評価指標として ImageNet2012 の validation データセットの上位 1 位認識結果を用いた．ノイズとブラーの精度は前述のσ_nとσ_bの 5 つすべてのパラメータにおける認識精度の平均を示している．各列で最も高い認識精度を示しているものを太字で示しているが，提案手法が歪み無し画像，ノイズ画像，ブラー画像のいずれにおいても最も高い精度を示している．これは，中間制約損失を用いて，Fine-tuning の前後で DNN の深層のニューロンの反応パターンを一貫したものにすることで学習の効率化が図れる，という本研究の仮説を支持しているものと考えられる．MSE 損失を用いた手法はノイズ画像・ブラー画像について，従来の Fine-tuning よりも高い精度を示しているが，提案手法よりも低い精度となっている．これは，MSE を用いた中間層への制約でも Fine-tuning と比較すると学習効率を高める事が出来る事を示唆している．しかし，提案手法の歪み付き画像認識精度が MSE よりも優れていることから，MSE のようにユークリッド距離で類似するように制約するよりも，中間層のニューロンが反応するか否かのパターンを明示的に制限する提案手法がより有効である事を示していると考えられる．

5. まとめ

本研究では，Fine-tuning による反応パターンの変更を防ぎ，効率的な学習を実現するため，中間制約損失を用いた歪み付き画像認識手法を提案した．中間制約損失はシグモイド関数を用いた二値分類タスクとして定式化でき，Pytorch や TensorFlow などの標準的な DNN フレームワーク上で実装が可能である．

中間制約損失を用いて Fine-tuning を行った DNN は中間制約を用いない Fine-tuning 手法，及び平均二乗誤差損失を用いて中間出力を制限する Fine-tuning と比較して高い認識精度を示した．これは DNN の深層の反応パターンを適切に制御する事で学習効率を高める，提案手法の有効性を示唆しているものと考えられる．

今後の研究計画として，本研究で提案した中間制約損失の更なる精度向上を図るための新規技術の考案と中間制約損失を用いて Fine-tuning した DNN モデルの解析が挙げられる．

参考文献

[1] Dodge, S., Karam, L., “Understanding how image quality affects deep neural networks”, International Conference on Quality of Multimedia Experience (QoMEX). (2016)
[2] Vasiljevic, I., Chakrabarti, A., Shakhnarovich, G., “Examining the impact of blur on recognition by convolutional networks”, arXiv preprint, arXiv:1611.05760 (2016)
[3] Zhou, Y., Song, S., Cheung, N., “On classification of distorted images with deep convolutional neural networks”, International Conference on Acoustics, Speech, and Signal Processing (ICASSP). (2017)
[4] Dodge, S., Karam, L., “Quality robust mixtures of deep neural networks”, IEEE Transactions on Image Processing 27 5553–5562 (2018)
[5] Hossain, M.T., Teng, S.W., Zhang, D., Lim, S., Lu, G., “Distortion robust image classification using deep convolutional neural network with discrete cosine transform”. International Conference on Image Processing (ICIP). (2019)
[6] Zeiler, M. D., Fergus, R., “Visualizing and understanding convolutional networks”, European Conference on Computer Vision (ECCV). (2014)