畳み込みニューラルネットワークの学習過程の可視化

Visualization of learning process for Convolution Neural Network

坂井 創一 竹中 要一 sakai soichi takenaka yoichi

関西大学 大学院総合情報学研究科

Graduate School of Informatics, Kansai University

Convolutional Neural Network (CNN) is an image classifier using deep neural network. However, it hardly gives the evidence why it classifies an image into a class. To solve this problem, some methods producing visual explanations has been proposed. Grad-CAM produces visual information for localized important regions for a class in an input image. As well as the visual explanations of classification, it is important to visualize of the learning process. The performance of CNN, such as accurate classification, is highly rely on the parameters. We convince the visualization of the learning process helps the parameter tuning. We propose a method that visualize the learning process of our method using MNIST dataset. The result shows the proposed method can visualize the learning process for every class for every epoch, whereas the usual method cannot.

1. はじめに

近年,深層学習を用いた画像認識は様々なタスクで優れた結果 を出している.深層学習の一種である畳み込みニューラルネット ワーク (Convolutional Neural Network : CNN)[LeCun 98] は画像認識や画像キャプション生成などのタスクに利用され, 既存手法の性能を大きく上回っている.

CNN では、多くの畳み込み層や全結合層、プーリング層等 複数種類の層を重ねることによって計算処理が行われている。 それゆえ、識別器の構造が複雑化し、識別根拠や識別理由を 人間が理解することが難しく、誤識別原因の調査に曖昧さが 残り、識別精度の向上に多くの労力を要している。これらの 問題を解決するために、SmoothGrad [Smilkov 17] や LIME [Ribeiro 16] など CNN が推論を行なった際にその判断根拠と して、推論結果に大きく影響を与えた箇所を提示できるような モデルを作る研究がなされている。

その中でも一般的によく使われている Grad-CAM [Selvaraju 17] は入力画像中で出力クラスに大きく影響を与 えた領域を可視化することで, CNN による出力が入力画像 のどの部分に注目して行なわれているのかを提示している. Grad-CAM は医療のように利用者に判断の根拠を説明する必 要がある分野に対応することが考えられてきた.

しかし,ニューラルネットワークの設計者にむけて,学習時 の学習率や畳み込み層のカーネルサイズ,ドロップアウトの割 合といったハイパラメータの調節に役立たせるために判断根拠 を分析する研究などは見られなかった.

本研究では、このようなハイパラメータの調節といった精度 を上げるための情報の一つとして学習過程に着目する.学習過 程の段階で、CNN が入力画像中の対象領域を正しく注目して いるかを見ることで、テストデータに対する正答率や損失関数 の値とは違った評価値を得ることができ、多角的にモデルの性 能を判断することが考えられる.そこで学習過程を可視化する ために Grad-CAM の応用を行なった.

本稿では学習過程の可視化に Grad-CAM を応用した際,可



図 1: 入力画像



図 2: クラス 'cat' の可視化 図 3: クラス 'dog' の可視化

視化が行われないパターンがある事を示し、そのようなパター ンに対して可視化を可能にする手法を提案する.

2. 先行研究

Grad-CAM は, 畳み込み層の最終層を使うことにより, あ るクラスにおける入力画像中の注目した領域を可視化する手法 である.図1をGrad-CAM に入力として与え, 'cat'のクラス の注目領域を可視化すると図2が出力される.また同じ入力 を用いて, 'dog'のクラスの注目領域を可視化すると図3が出 力される.

Grad-CAM は次の手順によって可視化を行なっている.

1. クラス c に対する各チャンネルの重要度をもとめる.

連絡先: 坂井創一, 関西大学大学院総合情報学研究科, k603014@kansai-u.ac.jp



図 4: 学習過程の可視化

k を層のチャンネル数, c を求める目的クラス, α をチャ ンネルの重要度, A を可視化したい畳み込み層の k チャ ンネルが出力する特徴マップ, y を目的クラスのスコアと する.

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k} \tag{1}$$

式 (1) では畳み込みレイヤの特徴マップ A_{ij}^k がクラス c の スコアに与える影響,すなわち y_c の勾配を計算する.そ してチャンネルの重要度の重み α_k^c を得るために平均化し ている.

2. クラス c に対するニューラルネットワークの注目度マップ を求める.

$$L^{c}_{Grad-CAM} = ReLU(\sum a^{c}_{k}A^{k})$$
⁽²⁾

式 (2) では各チャンネルの重要度と a_k^c とそのチャンネル が出力する特徴マップ A^k を足し合わせ, ReLU 関数を 掛けている.

3. 注目度マップを可視化するため,出力が1から0の間に なるよう正規化を行う.



図 5: 微分箇所

$$L^{c}_{Grad-CAM} = \frac{L^{c} - min(L^{c})}{max(L^{c}) - min(L^{c})}$$
(3)

式(3)によって可視化されたヒートマップが与えられる.

この手法を学習過程の可視化に応用した際,図3の入力に 図4(左a)に示すように注目箇所が可視化が行われなかった.

3. 提案手法

従来手法において可視化が行われなかった際,目的クラス のスコアである y^c が 1 なっていた. y^c の微分値を求める際, CNN の全結合部のシグモイド関数と softmax 層によって勾配 情報が失われ,チャンネルの重要度 α_k^c が 0 になってしまう問 題があることがわかった.

提案手法では、 α_k^c の値を0にしないため、チャンネルの重 要度の計算式に変更を加えた. n を層の数、x をパーセプトロ ン、 x^n をn 層目のパーセプトロンとする. 従来手法では式(1) の中で(図 5:a)の位置にある目的クラスのスコアの値 α_k^c に対 して行われていた微分を、提案手法では(図 5:b)の位置にあ る softmax 層に入力するユニットのシグモイド関数をかける 前の値 x_i^n に対して微分を行う.

つまり,提案手法では従来手法では式 (1) で計算を行なって いた α_k^c を代わりに式 (4) によって行う.

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial x_n^c}{\partial A_{ij}^k} \tag{4}$$

こうすることによって, softmax 層やシグモイド関数で勾配 が消失している場合にも注目領域の可視化が可能となる. 結 果,図4(右:b)のように可視化がされた.

4. 実験

手書き数字画像データセットである MNIST^{*1} を利用し,従 来手法と提案手法の比較実験を行なった.

実験に使用した CNN の構造は 1 である. 畳み込み層の活性 化関数には ReLU を, Batch size は 100 を採用した. 全結合 層のみに 0.5 の選出確率で Dropout を導入した. 1 エポック

^{*1} http://yann.lecun.com/exdb/mnist/

1 1 1 1 1 1 1 1 1 1	The 33rd Annual	Conference o	f the Japanese	Society for A	Artificial Intelligence,	2019
---------------------------------------	-----------------	--------------	----------------	---------------	--------------------------	------

表 1: CNN の構成												
		入力次元		出力次元								
層番号	種類	Η	W	D	Η	W	D	カーネル	ストライド			
1	畳み込み	28	28	1	26	26	32	3×3	1×1			
2	畳み込み	26	26	32	24	24	64	3×3	1×1			
3	プーリング	24	24	64	12	12	64	2×2	2×2			
4	Flatten	1	1	9216	1	1	128	-	-			
5	Flatten	1	1	128	1	1	10	-	-			





図 6: 実験: 従来手法

に学習データ 60000 枚中からランダムで 10000 枚選んで学習 を行なった.テストデータから 0 から 9 までの数字を 1 枚ず つ選び,1 エポック毎に従来手法と提案手法の Grad-CAM を 掛け,可視化が行われていない画像の枚数を調べた.それぞれ 実験を行なった,従来手法の出力結果を図 6 提案手法の出力 結果を図 7 に示す.結果から従来手法では可視化されなかった 画像においても,可視化されていることが分かる.学習を 100 エポック行った結果,可視化が行われていなかった画像は 500 枚のうち,従来手法では 217 枚,提案手法では 0 枚であった.

5. 終わりに

提案手法により, CNN の学習過程の可視化が可能となるこ とを示すことができた.指定したクラスの判断根拠を注目領 域としてエポック毎に出力が可能である,これを用いることに より

モデルのあるクラスにおける判断根拠が誤っていた場合,他 クラスの判別にどのように影響を与えるかや,誤った判断根拠 にどう修正を施すか,判断根拠を元にどのような追加学習を行 うことが適切かなどを検証する手段として利用できる事を示 した.

図 7: 実験:提案手法

参考文献

- [LeCun 98] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. Proc. of the IEEE, pages 2278-2324, 1998.
- [Selvaraju 17] Selvaraju, Ramprasaath R., et al. :Gradcam: Visual explanations from deep networks via gradient-based localization. (2017) IEEE International Conference on Computer Vision (ICCV). IEEE, 2017.
- [Smilkov 17] D. Smilkov, N. Thorat, B. Kim, F. Vigas, and M. Wattenberg. Smoothgrad: removing noise by adding noise. arXiv preprint arXiv:1706.03825, 2017.
- [Ribeiro 16] Ribeiro, Marco Tulio, Sameer Singh, and Carlos Guestrin. "Why should I trust you?: Explaining the predictions of any classifier." Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. ACM, 2016.