

複数の注視領域に基づく色恒常性

上坊瑞希[†]
東京電機大学[†]

小篠裕子[‡]
東京電機大学[†]

1 はじめに

照明光や環境光は画像内の物体の色に大きく影響するため、画像における色恒常性 (Color Constancy) に関する研究が議論されている。色恒常性とは、照明条件の変化に左右されずに物体の真の色を認識できる色覚特性を指す。近年では、Convolutional Neural Network (CNN) の導入により色恒常性の研究は大いに進歩した。

CNN を用いた関連研究の1つである FC4 [1] は、視覚的説明を光源色推定に導入した手法である。ここで、視覚的説明とは、画像内のどの領域が、与えられたクラスラベルへの識別結果をもたらしたかを、勾配情報に基づいて可視化することで、深層学習モデルの判断根拠を解析する手法である。

FC4 による光源色推定は、視覚的説明により推論した特定の領域のみに依存してしまうという問題がある。

そこで本稿では、単独の深層学習モデルの判断根拠に捉われない光源色推定をする手法を提案する。実験にて、提案手法の有効性はある程度はみられたが、未だ改善の余地があることが分かった。

2 従来研究

画像における色恒常性タスクでは、光源色推定に関する研究が多い。光源色推定とは、RGB 画像 $I \in \mathbb{R}^{H \times W \times 3}$ が入力として与えられたときに、画像 I に投影されている光源色 $p_g = (r, g, b)$ を正規化した $\hat{p}_g = \frac{p_g}{\|p_g\|_2}$ を推定することである。

Hu ら [1] が提案した光源色推定手法、FC4 [1] がある。FC4 は、入力画像から切り出された小領域 (パッチ) に、推定の確信度を表すスコア (Confidence map) を割り当てることで光源色を推定する CNN アーキテクチャである。Confidence map は、ネットワークが推論時に注視した領域をヒートマップで可視化した Attention

map と類似しているため、本手法は視覚的説明手法に基づいているとも言える。

3 提案手法

本稿では、単独の深層学習モデルの判断根拠に捉われない光源色推定をする手法を提案する。具体的には、1枚の入力画像から複数の Confidence map を得るためのアーキテクチャと、得られた Confidence map が一致しないよう学習させるための損失関数を提案する。

提案するアーキテクチャの全体図を図1に示す。アーキテクチャは n 個の FC4 [1] のネットワークにより構成される。以降1つの FC4 のネットワークをヘッドと呼ぶ。 k 番目のヘッド k から得られる p^k は、入力画像内のパッチから推定された照明色 \hat{p}^k と Confidence map c^k を掛け合わせたものである。 p^k を n 個連結させ、Confidence map の特徴をより重要視させるための畳み込み処理をし、正規化したものを推定色として算出する。

提案する損失関数を以下の式 (1) のように定義する。

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{illum}} + \lambda \mathcal{L}_{\text{conf}}. \quad (1)$$

ここで、 $\mathcal{L}_{\text{illum}}$ は色推定の誤差を、 $\mathcal{L}_{\text{conf}}$ は Confidence map についての誤差を表している。以下に、それぞれの式の詳細を述べる。

$\mathcal{L}_{\text{illum}}$ は真値の光源色 L と推定された光源色 \hat{L} を用いて算出される角度誤差であり、色恒常性のアルゴリズムで最も広く用いられている [2]。

$$\mathcal{L}_{\text{illum}} = \arccos\left(\frac{L \cdot \hat{L}}{\|L\| \|\hat{L}\|}\right). \quad (2)$$

$\mathcal{L}_{\text{conf}}$ は式 (3) に示すように、大きな値を示すピクセル領域を可能な限り分離することができる [3]。ここで、 c_{xy}^i と c_{xy}^j は2枚の Confidence map c^i と c^j の (x, y) 番目の画素を表している。

$$\mathcal{L}_{\text{conf}} = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \left(2 \cdot \frac{\sum_{xy} \min(c_{xy}^i, c_{xy}^j)}{\sum_{xy} c_{xy}^i + c_{xy}^j} \right). \quad (3)$$

Color Constancy with Multiple Attentions

[†]Mizuki Uebo, Graduate School of System Design and Technology, Tokyo Denki University

[‡]Yuko Ozasa, Graduate School of System Design and Technology, Tokyo Denki University

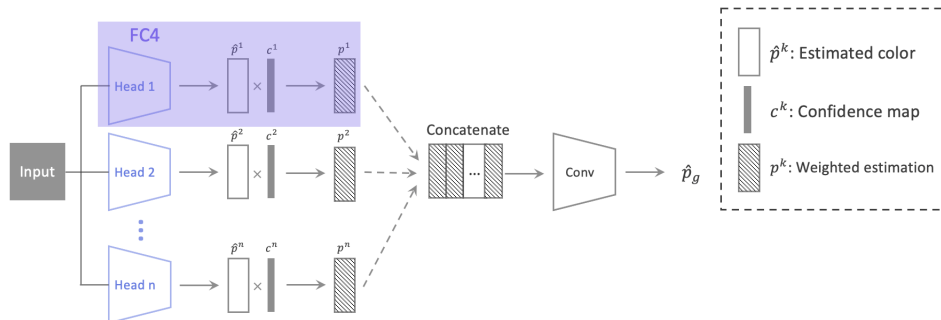


図 1: 提案アーキテクチャの構造.

表 1: 角度誤差の比較 (度).

Dataset	Methods	Mean	Median	Tri-mean	Best25%	Worst25%
Gehler [4]	FC4 [1]	1.65	1.18	1.27	0.38	3.78
	Ours	2.53	1.77	2.03	0.59	5.74
Cube+ [5]	FC4 [1]	1.35	0.93	1.01	1.35	3.24
	Ours	2.41	1.29	1.48	0.37	6.43

4 実験

4.1 実験条件

複数のヘッドで構成された提案モデルの性能を評価するため、ヘッドが1つである単独モデルのFC4と提案手法を比較する。実験では、FC4と同様に単一の照明光を推定することを想定し、提案モデルのヘッドの数を $n = 2$ とした。

データセットには Color Checker Dataset(CCD) [4] と Cube+ [5] を使用し、従来研究と同様に 3-fold cross validation を採用した。CCD [4] はデータセットの総数 568 枚を、Cube+ [5] はデータセットの総数 1707 枚の中からランダムに取り出された 570 枚を使用した。最適化手法として Adam を使用し、学習率は 3×10^{-4} 、エポック数を 1000、バッチサイズを 16、 $\lambda = 1 \times 10^{-4}$ とした。従来研究に従い、平均値、中央値、外れ値を除いた平均値、誤差の最小 25% の平均値、誤差の最大 25% の平均値の 5 つの指標を採用する。

4.2 結果

表 1 に従来手法と提案手法の性能評価の比較を示す。各数値は角度誤差であり、数が小さければ小さいほど性能が良いことを表している。次に、提案手法により得られた Confidence map を可視化したものを図 2 に示す。同図左から入力画像、FC4 [1] より得られた Confidence map, (a) および (b) は提案手法の各ヘッドから得られた Confidence map, 推定結果, 真値の画像を示している。提案手法では、ヘッド間で注視領域が異なること

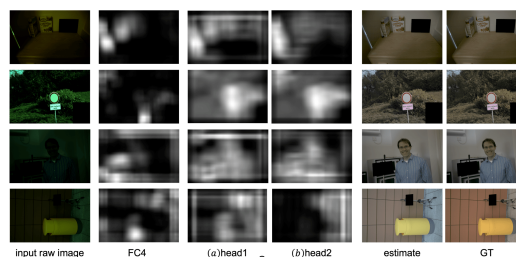


図 2: Confidence map と推定結果の可視化.

が確認できる。一方で最下段は検証データの誤差が大きい結果の例である。ヘッド間で Confidence map 内に大きな差が生じると、色推定に寄与しない領域を注視したために精度が低下したと考えられる。

5 まとめ

本稿では、複数の Confidence map を得るためのアーキテクチャを提案し、さらに得られた Confidence map を引き離す損失関数を提案した。データの種類の提案手法に与える影響について調査することを今後の課題とする。

参考文献

- [1] Hu, Y. et al.: FC4: Fully Convolutional Color Constancy with Confidence-Weighted Pooling, *Proc. of CVPR* (2017).
- [2] Barron, J. T.: Convolutional Color Constancy, *CoRR* (2015).
- [3] Liu, W. et al.: Towards Visually Explaining Variational Autoencoders, *Proc. of CVPR* (2020).
- [4] Gehler, P. V. et al.: Bayesian color constancy revisited, *Proc. of CVPR* (2008).
- [5] Banić, N. et al.: Unsupervised Learning for Color Constancy, *CoRR*, Vol. abs/1712.00436 (2017).

本研究を進めるにあたり、ご協力いただいた慶應義塾大学 (現所属:KDDI 総合研究所) 鶏内朋也様に深く感謝いたします。