

写真撮影スキル向上のための VR トレーニングシステム

小林 大貴[†] 長尾 確[‡]名古屋大学 大学院情報学研究科^{†‡}

1. はじめに

近年, SNS の普及により Web 上への写真の投稿が増加している. 従来写真とは, 主に記録のために用いられるツールであった. しかし SNS の普及に伴い, 写真に対する考え方に変化が生じ, 写真を共有する文化が生まれた. その根拠として 2017 年の流行語大賞には“インスタ映え”が選ばれた. このように, 写真は自分だけが見るものではなく, 誰かに見てもらうものという側面が強くなったと言える. これにより, 写真を上手く撮りたいという願望が強くなったと考える. また近年, スマートフォンのカメラ性能が向上しており, 画質や処理能力が高まっただけでなく, F 値 (絞り) やレンズの切り替えなどの調整も行えるようになった. このため写真撮影の表現の幅が大きく広がったと言える.

カメラがより身近になった一方で, カメラに関心を寄せる人が抱える課題もある. それは被写体となってくれる人や物を探すことは難しくなかなか練習できないこと, アドバイスしてくれる人が身の回りにいないことが挙げられる. そこで本研究では, 利用者が仮想世界内で撮影した写真に対して, 評価・フィードバックを行う VR Photo トレーニングシステムを提案する.

2. 撮影した写真の評価方法とデータセットの作成

本システムの構成を図 1 に示す. 本研究では, 仮想世界内で, 静止したアバターの撮影を行い, リアルタイムでその写真の評価を行う. 写真を評価する枠組みとして, 畳み込みニューラルネットワーク (CNN: Convolutional Neural Network) を利用した美的評価モデルを用いる.

2.1 写真の美的評価モデル

美的評価モデルとして, 精度と計算コストの観点から NIMA (Neural Image Assessment) モデル[1]を用いた. そのアーキテクチャを図 2 に示す. 先行研究では, baseline のモデルとして MobileNet, VGG16, InceptionV2 が用いられてい



図 1: システムの構成図

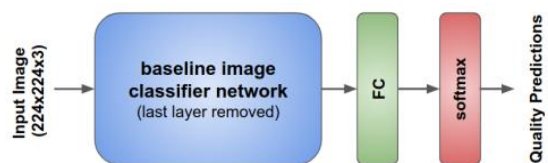


図 2: NIMA モデル

た. また NIMA モデル[1]では, 写真の良し悪しを 2 値分類で予測するのではなく, スコア分布を予測するため, 損失関数として以下の EMD (Earth Mover's Distance) を用いる.

$$EMD(p, \hat{p}) = \left(\frac{1}{N} \sum_{k=1}^N |CDF_p(k) - CDF_{\hat{p}}(k)|^r \right)^{\frac{1}{r}} \quad (1)$$

EMD は 2 つの確率分布間の距離を測る尺度である. 左辺において, p, \hat{p} はそれぞれ正解スコア, 予測スコアの確率分布を表す. 右辺において, N は順序付きクラス数, CDF は累積分布関数, r はノルムを表す.

2.2 データの収集

本研究では, アプローチとして, 評価精度の向上を目指すため Fine tuning を行う. そのため, まずはデータ収集を行った. 以下の手順で, 仮想世界内での写真の自動撮影を行った.

1. アバターとポーズをランダムに選択し, そのアバター (被写体) を任意の場所に生成
2. アバターからの距離・角度・高さをランダムに取得し, その場所にカメラを移動
3. レンズ (焦点距離) ・ F 値 (絞り) ・明るさの値をランダムに選択

VR Training System for Improving Photography Skills
[†]KOBAYASHI, Hiroki (kobayashi@nagao.nuie.nagoya-u.ac.jp)
[‡]NAGAO, Katashi (nagao@i.nagoya-u.ac.jp)
^{†‡}Graduate School of Informatics, Nagoya University

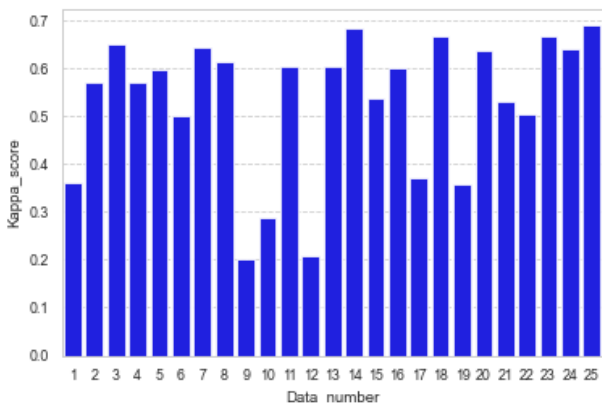


図 3: 5 段階評価におけるカッパ係数の値

4. アバターがカメラの方向を向いている写真と向いていない写真を撮影
5. 生成したアバターの削除

この手順で写真データを 5000 枚生成した。

2.3 正解データの作成

生成した写真データ 5000 枚に対して、美的評価のラベル付けを行った。評価者はインスタグラムの投稿数が 10 枚以上の投稿者（以下、協力者）18 人および自分である。ラベル付けは 5 段階評価と 2 段階評価の 2 種類とした。機械学習では、5 段階評価のデータを使用した。

また評価データの収集方法として、Google フォームと Google フォトを用いた。5000 枚の写真をランダムに 25 分割し、協力者は基本的に 200 枚の写真を評価、自分は 5000 枚の写真を評価し、1 枚の写真に対して 2 名で評価した。

2 名の評価の一致度を表す基準として、以下の Cohen の重み付きカッパ係数[2]を用いた。

$$\kappa = \frac{p_o(w) - p_e(w)}{1 - p_e(w)} \quad (2)$$

(2)式において、 $p_o(w)$ は 2 名の評価がどの程度類似しているかを表す確率変数、 $p_e(w)$ は 2 名の評価がどの程度偶然に類似してしまうかを表す確率変数である。また 5 段階評価における重み付きカッパ係数の値を図 3 に示す。本研究では、カッパ係数の値が 0.4 より大きい 3800 枚のデータを使用した。

3. 実験結果

NIMA モデル[1]の事前学習では、AVA データセット[3]を用いて、パラメータの学習を行った。そして収集した 3800 枚のデータに対して、train データと test データの比率を 9:1 に分割し、Fine tuning を行った。また NIMA モデル[1]の baseline モデルとして MobileNet と InceptionV2 を試した。ここでは精度がより良かった MobileNet

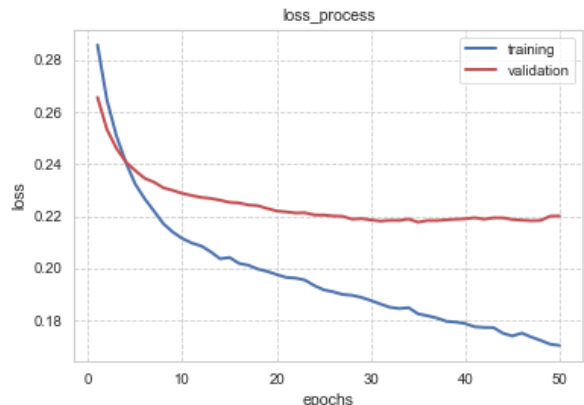


図 4: 損失関数の推移

表 1: Fine tuning 前と Fine tuning 後の精度

	accuracy	f1_score
before	0.605263	0.659091
after	0.821053	0.832512

の損失関数の推移と精度をそれぞれ図 4、表 1 に示す。この結果より、Fine tuning を行うことで精度の向上が確認された。

4. まとめと今後の課題

本研究では、仮想世界内においてアバターの写真を撮影し、美的評価モデルにより評価・フィードバックを行うことで、写真撮影スキルの向上を支援する VR Photo トレーニングシステムの提案を行った。

今後の課題としては、写真評価のより具体的なフィードバックを行うことが挙げられる。写真の良し悪しを提示するだけでなく、“何が良くて何が悪かったか”などについてもフィードバックできると良いだろう。そのため、写真の総合的な評価を行うだけでなく、タスクを分割して構図・色味・ボケ具合など個々の評価も行っていく。今回のデータ収集では、写真の良し悪しを総合的に評価してもらったが、次回のデータ収集の際には、構図等の個々の評価も行ってもらいたいと考えている。さらに将来的には、構図推薦などの機能も実現する予定である。

参考文献

- [1] H. Talebi and P. Milanfar. Nima: Neural Image Assessment. TIP, 2018. 3.
- [2] HL. Kundel and M. Polansky. Measurement of Observer Agreement. Radiology 2003; 228:303–308.
- [3] N. Murray, L. Marchesotti, and F. Perronnin. AVA: A Large-Scale Database for Aesthetic Visual Analysis. In CVPR, 2012. 3.