

3DCG 画像を教師データとした Deeppose による人物姿勢推定

宮森勇作^{†1} 川村秀憲^{†1}

概要 : 人物姿勢推定の研究では、多層の畳み込みニューラルネットワークを用いることで最高水準の姿勢推定を実現している。畳み込みニューラルネットワークの学習には、ラベル付けされた膨大な数のデータセットが必要である。しかし、現実世界で膨大な数のシチュエーションの人物画像を用意するのは難しい。より多くのシチュエーションの人物画像と姿勢情報を同時に自動生成する仕組みを作り、生成されたデータセットを教師データとして機械学習を行うことでより正確な人物姿勢推定が可能であると考えられる。本研究では、学習に十分な数のデータセットを用意するために、3DCG を用いて人物画像と姿勢情報を自動生成した。生成したデータセットを教師データとして深層学習を行った。また、現実の人物画像に則した姿勢推定モデルとなるか否かを検証する前段階として 3DCG の人物姿勢の推定を行い、精度検証を行った。

キーワード : 人物姿勢推定, 3DCG 画像

1. はじめに

人物姿勢推定の研究では、深層学習と呼ばれる多層の畳み込みニューラルネットワークを用いることで最高水準の姿勢推定を実現している。人物姿勢推定を目的とした畳み込みニューラルネットワークの学習には、人物画像と人物姿勢情報を組とした膨大な数の教師付きデータが必要なため、学習に十分な教師データを用意することは難しい。したがって、多様多様な教師データを自動生成する仕組みを作り、生成されたデータセットを教師データとすることで効率的に学習を行うことができ、より正確な人物姿勢推定が可能になると考えられる。

本稿では、以下のような構成で議論を進める。第 2 章で深層学習を用いた人物姿勢推定の関連研究を例に挙げる。第 3 章では 3DCG を用いることで教師データを自動生成するツールを作成し、学習に必要なデータセットを生成する。第 4 章では人物姿勢推定の手法である Deeppose を用いて、第 3 章で生成したデータセットを教師データとして学習を行い、姿勢推定の結果をまとめる。第 5 章では今後の課題と展望についてまとめる。

2. 関連研究

深層学習を用いた人物姿勢推定の研究として Deeppose[Toshev 2014]がある。Deeppose は深層学習を用いた人物姿勢推定を行った最初の研究であり、姿勢推定タスクを回帰問題としたアプローチをとっている。

深層学習のネットワーク構造として AlexNet[Krizhevsky 2012]と ResNet[He 2015]等がある。これらは ILSVRC と呼ばれる画像認識のコンペティションにおいてトップクラスの成績を収め、汎用性の高いネットワークであるため多くの研究で利用されている。

本研究では、人物姿勢推定の手法として Deeppose、ネッ

トワーク構造として AlexNet と ResNet を用いて、学習を行う。学習に用いるデータセットは 3DCG 画像の自動生成ツールを作成することで十分な数を用意する。

3. 3DCG 画像データセットの自動生成

Deeppose の学習を行うためには人物画像と人物姿勢情報を組としたデータセットが必要である。本研究では、3DCG を用いてデータセットを自動生成するツールを作成し、姿勢の異なる 10000 組の教師データを生成した。人物の 3D モデルには比較的利用しやすい 3D モデルである UNITY-CHAN を用いた[4]。



図 1 生成したデータセットから無作為に抽出した画像群
また、人物の全身姿勢の推定を目標として画像上の頭や手足などの全身の関節 16 点の座標を姿勢情報とした。



図 2 元画像(左)と生成された姿勢情報を描画した画像(右)

^{†1} 北海道大学大学院情報科学研究科
Hokkaido University

4. Deeppose による学習

3章で生成したデータセット10000枚のうち9000枚を訓練データ、1000枚をテストデータとして、Deepposeの手法に基づいて学習を行った。学習時のネットワーク構造にはAlexNetとResNetを用いた。また、活性化関数にはAdam、AdaGrad、MomentumSGDを用いた。以下の図3は横軸をエポック数、縦軸を誤差として学習の経過をグラフにしたものである。

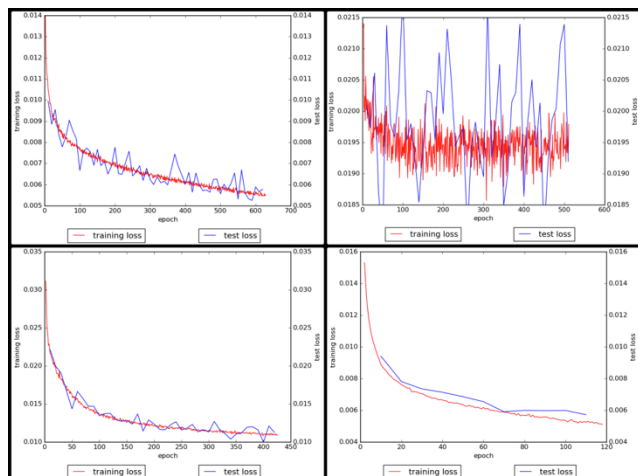


図3 AlexNet-AdaGrad (左上)・AlexNet-Adam (右上)
AlexNet-MomentumSGD (左下)・ResNet-AdaGrad (右下)

グラフの比較によって、活性化関数にAdamを使った場合は発散し、それ以外は学習が進んでいることがわかる。また、活性化関数にAdaGradを用いた場合、ResNetはAlexNetに比べて少ないエポック数で学習が進んでいることがわかる。

さらに、学習したモデルを用いて姿勢推定を行った。以下の図4は教師データの画像群と活性化関数にAdaGrad、ネットワークにAlexNetを用いて学習したモデルにより姿勢推定を行った画像群である。

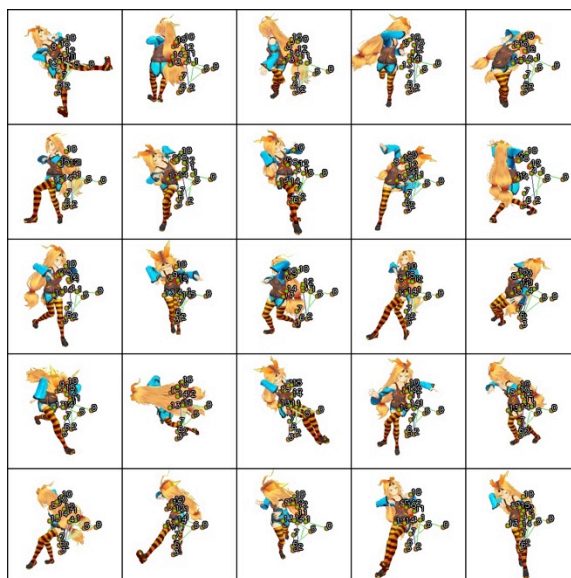
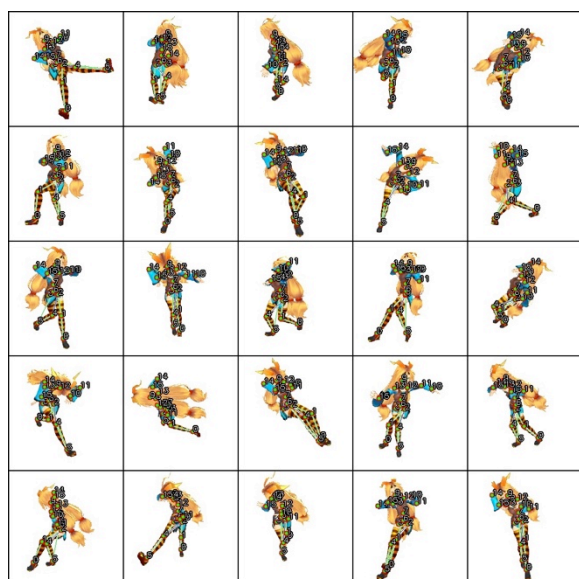


図4 教師データの画像群(上)、推定データの画像群(下)

推定結果からわかるように3DCG画像を教師データに用いた学習モデルは推定精度が低い結果となった。

5. 今後の課題と展望

本研究では、学習に十分な数のデータセットを用意するために、3DCGを用いて人物画像と姿勢情報を自動生成した。生成したデータセットを教師データとして多層の畳み込みニューラルネットワークの学習を行った。さらに、現実の人物画像に則した姿勢推定モデルとなるか否かを検証する前段階として3DCGの人物姿勢の推定を行った。結果として、推定した姿勢は実際の姿勢と大きく異なった。原因としては、学習に用いた画像の解像度の低さ、3Dモデルの形状により全身が隠れた状態のデータが複数存在したこと、データセットのシチュエーションの少なさなどが考えられる。

今後は、学習に用いる画像の解像度を大きくすることで推定精度が向上するか否かを検証する。また、実写に近い人物モデルを利用して画像を生成し、データセットをより多くのシチュエーションで生成できるようなツールを作ることによって現実の人物についても正確な姿勢推定を目指す。

参考文献

- [1] Toshev, Alexander, and Christian Szegedy. "Deeppose: Human pose estimation via deep neural networks." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014.
- [2] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems. 2012.
- [3] He, Kaiming, et al. "Deep residual learning for image recognition." arXiv preprint arXiv:1512.03385. 2015.
- [4]"UNITY-CHAN OFFICIAL WEBSITE".
<http://unity-chan.com/contents/guideline/>.