

CNN を用いた光沢感の要因の解析

弘部 大知[†] 青山 俊弘[†]

鈴鹿工業高等専門学校電子情報工学科[†]

1. はじめに

近年の AI 技術の発達により AI が画像や映像中の物体が何であるかを識別する能力は著しく向上した。一方で今後は物体の品質や状態がどうであるか、といったより抽象的な識別能力が重要となることが予想され、そのためには物体表面の質感認識が必要であると考えられる。そこで本研究では質感の一つである光沢感に注目し、様々なデータセットを学習させた CNN がそれぞれどのようにして光沢感の認識を含め表面材質を識別しているのかを解析した。先行研究として 3 段階までの浅い CNN を光沢感のある材質の識別に特化する形で学習を行わせた研究がある[1]が、本研究では VGG16[2]をベースとしたより深い CNN をオートエンコーダ(以下 AE)を基にして作成することで、光沢感の認識に特化したものではなく人間の視覚野における活動に近いものとなるようにした。このような CNN がどのようにして光沢感を認識しているのかを解析することで人間の脳内でどのようにして光沢感が認識されているのかといった点について考察を試みる。

2. 研究概要

CG を用いて 3D モデルの画像群を用意し、それを 3D モデルに適用されている材質に応じて分類するというタスクを行う CNN(以下タスク CNN)を作成する。タスク CNN の畳み込み処理の部分はオートエンコーダ(以下 AE)によって事前学習を行っている。AE は自然画像を使用して学習を行ったものである。タスク CNN における AE での事前学習を行った部分に関しては、学習中には重みを更新しない。そうすることにより材質を分類する際に使用する特徴量を AE が学習した特徴量を組み合わせる形で学習することになる。

このようにすることで、一から材質の分類のために学習を行った場合よりも人間が視覚情報を用いて材質を識別するプロセスに近いものになることが期待できる。

使用する材質としては光沢のある材質を 4 種類、光沢の無い材質を 4 種類の合計 8 種類を用意した。これらを用いて光沢のある材質間での 4 クラス分類や、光沢のある材質と光沢の無い材質の 2 クラス分類といったタスクを行いその分類精度を調べる。また、Activation Maximization[3]というネットワーク中の特定のユニットの出力ができるだけ大きくなるような入力画像を生成する手法により各材質の判断根拠となる特徴を観察する。

3. データセットの作成

3.1 AE のデータセット

AE の学習のために使用する自然画像のデータセットは ImageNet の画像を用いて作成した。ImageNet に含まれる画像の中から被写体の位置を示すバウンディングボックスが 256×256 に収まる画像を選び、それらをバウンディングボックスを中心として 256×256 に切り取った部分画像を学習用に 65099 枚、評価用に 10000 枚作成し AE の学習に使用した。



図 1: AE の学習データセットの一例

3.2 タスク CNN のデータセット

タスク CNN の学習に使用する CG 画像は Unity 2019.1.0f2 を使用して生成した。Standard シェーダを適用した 3D モデルと DirectionalLight をシーンに配置し、3D モデルには背景画像をもとに環境光を適用し、材質に応じて背景画像を映り込ませる。光沢のある材質 4 種類はそれぞ

Generational mechanism of glossiness in CNN

[†] Daichi Hirobe and Toshihiro Aoyama. National Institute of Technology, Suzuka college

れ Standard シェーダーの `_Metallic` と `_Glossiness` の値をそれぞれ(0.8, 0.5)、(0.2, 0.8)、(0.5, 0.9)、(1, 0.6)としたものである(値は(`_Metallic`, `_Glossiness`)の順)。光沢感の無い材質 4 種類はそれぞれ(0, 0)、(0.5, 0)、(1, 0)、(0.6, 0.2)としたものである。

Skybox に使用する実写画像は学習データ用に 22 枚、評価データ用に 4 枚、テストデータ用に 4 枚それぞれ異なるものを用意した。各背景画像につき `DirectionalLight` の照射角と背景画像の映り込む部分が異なる 189 枚の画像を生成した。これにより学習データとして $22 \times 189 \times 8 = 33264$ 枚、評価データとして $4 \times 189 \times 8 = 6048$ 枚、テストデータとして $4 \times 189 \times 8 = 6048$ 枚の画像を用意することができた。全ての画像は大きさ 256×256 の jpg 画像である。

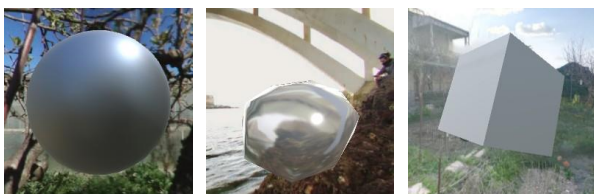


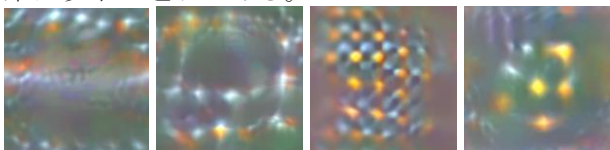
図 2: タスク CNN の学習データセットの一例

4. 結果

いくつか行った実験の中から抜粋して結果を示す。

4.1 球体の 3D モデルのみを用いた光沢のある材質の 4 クラス分類結果

分類精度 98.4% という非常に高い精度で分類することができた。Activation Maximization の結果は以下の通りである。



Motoyoshi らが提案した光沢感の識別モデル[4]によれば光沢感の強さはオン中心フィルタの出力によって決定されるが、それと合致するように各材質の特徴として明点を多く含む結果が得られた。

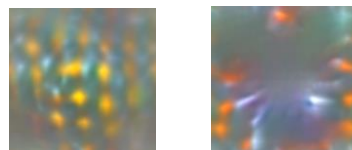
4.2 複数種類の形状の 3D モデルを用いた光沢のある材質の 4 クラス分類結果

球体の 3D モデルのみを使用した場合と比べて分類精度は 67.9% に落ちた。これは 3D モデルの種類を増やしたことにより形状未知の状態となったためだろう。人間が行う質感認識は物体の

形状認識と複合して行われていることが[5]によって示されているがタスク CNN ではそこまでの学習を行うことは困難なようだ。

4.3 複数種類の形状の 3D モデルを用いた光沢のある材質と光沢の無い材質の 2 クラス分類結果

光沢のある材質と光沢の無い材質は視覚的に大きく異なり、そのため形状未知の状態でも分類精度 81.7% と高い値を得ることができた。Activation Maximization の結果は以下の通りである。



左が光沢のある材質の結果であり右が光沢の無い材質の結果である。Motoyoshi らが提案した光沢感の認識モデルによれば光沢感のない材質はオフ中心フィルタの出力によるはずだがそのような特徴は見られない。

5. まとめ

本研究では光沢のある物体、無い物体の画像を用いて、CNN に材質による質感認識を行わせた。結果、類似の材質認識は形状既知の状態であれば十分行うことができるが形状未知の状態では難しいことが分かった。また光沢のある材質の識別にはオン中心フィルタの影響が大いに見られた一方で光沢の無い材質の識別にはオフ中心フィルタの影響を示す結果は見られなかった。

参考文献

- [1] 重川裕和, 岡谷豊之, ディープニューラルネットワークを用いた物体表面の光沢感認識. 2013
- [2] Simonyan, Karen and Zisserman, Andrew. Very deep convolutional networks for large-scale image recognition. 2014
- [3] Erhan, Dumitru and Bengio, Yoshua and Courville, Aaron and Vincent, Pascal. Visualizing higher-layer features of a deep network. 2009
- [4] Motoyoshi, Isamu and Nishida, Shin'ya and Sharan, Lavanya and Adelson, Edward H. Image statistics and the perception of surface qualities. 2007
- [5] Marlow, Phillip J and Anderson, Barton L. Material properties derived from three-dimensional shape representations. 2015