2M4-OS-32a-4

設計支援に向けたアフォーダンスを誘発する特徴の識別

Identification of Discriminative Features Affording a Particular Action using an Analysis of CNN

*1中田　勇介　　*1荒井　幸代
Yusuke Nakata　　Sachiyo Arai

*1千葉大学大学院融合理工学府　地球環境科学専攻　都市環境システムコース
Department of Urban Environment System, Division of Earth and Environmental Sciences,
Graduate School of Science and Engineering, Chiba University

Ideal products offer proper usages to users intuitively. Affordance is the qualities of an environment(e.g. product) that define its possible uses. Improvement of product's affordance leads to load reduction and enhanced safety. In this paper, we think that features of product(e.g. color, shape) induce affordance. Designer's experiences are required for making product to have proper affordance, because knowledge on features which induce affordance are implicit knowledge. Extraction of this knowledge makes it possible for all designers to enhance product's affordance. In this paper, we propose a method that identify discriminative features affording particular action with dataset of image of product and affordance user perceived. In more detail, we train a CNN model to predict product's "sittable" affordance, and analyze trained CNN to identify discriminative features which induce sittable affordance.

1. はじめに

理想的な設計物は，本来の使い方を直感的に利用者に伝える．設計物などの環境が動物に提供する「行為の可能性」を意味する概念はアフォーダンス [Norman 99] といわれ，設計物が備えるアフォーダンスと利用目的の合致は，利用者の負荷軽減や誤用による危険防止につながる．本研究では，設計物が備える色や形などの特徴がアフォーダンスを誘発すると考え，その特徴の識別を目的とする．従来，設計物が備えるべき色や形などの「アフォーダンスを誘発する特徴」は，設計者の経験や勘に内在する暗黙知であった．しかし，この暗黙知を形式化することができれば，経験の有無に関わらず，利用目的に沿ったアフォーダンスを設計物に持たせることが可能になる．

そこで，本研究では,「座る」という利用目的の設計物が誘発すべき「座れる」アフォーダンスを例として，設計物の持つべき特徴を識別する方法を提案する．提案法は 2 段階からなる．はじめに，設計物の画像データ群を畳み込みニューラルネットワーク（CNN）の入力，利用者（被験者）が感じる「座れる／座れない」の知覚を教師データとして「座れる」アフォーダンスを学習させる．次に，座れるアフォーダンスを誘発する特徴を識別するための方法を四つ取り上げ，それぞれの方法によって識別される「座れる」アフォーダンスを誘発する特徴を比較し考察する．

2. 対象問題のモデリング

人間は状態 s とアフォーダンス Q を知覚し行動 a を決定する．ここで，任意の行動 a のアフォーダンスを Q_a と表す．また，特徴ベクトル $\phi : s \mapsto [0,1]^k$ を定義し，特徴ベクトル ϕ を構成する要素を特徴 λ とよぶ．本研究では，状態 s と座れるアフォーダンス Q_a を所与とし，座れるアフォーダンス Q_a を誘発する特徴 λ を識別する．

連絡先: 中田勇介，千葉大学大学院融合理工学府　地球環境科学専攻　都市環境システムコース，千葉市稲毛区弥生町 1-33，a1019nakata@gmail.com

2.1 畳み込みニューラルネットワーク

本研究では，画像データに対して高い学習能力を有する機械学習モデルである畳み込みニューラルネットワーク (以下 CNN と表記)[LeCun 89] に，座れるアフォーダンス Q_a を学習させる．Figure 1 に座れるアフォーダンス Q_a を学習する CNN の概略を示す．

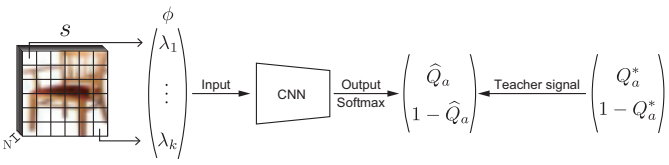


Figure 1: Learn affordance with CNN

ここで，N は学習に用いる画像データの数を，$\widehat{Q}_a$, Q_a^* はそれぞれ座れるアフォーダンス Q_a の推定値，教師データを指す．

2.2 CNN の解析法

CNN の入力層，または中間層の出力が CNN の出力に与える影響を可視化する手法に Saliency map[Simonyan 13], Guided Saliency map[Springenberg 14], Grad-CAM[Selvaraju 16], Guided Grad-CAM[Selvaraju 16] がある．以下でこれら四つの手法について説明する．

■ Saliency map

Saliency map は，入力画像の各ピクセルが CNN の出力に及ぼす影響の大きさを可視化する手法である．いま，ある画像 s_0, 出力クラス a, 画像に対する CNN の出力 $Q_a(\phi_0)$ が与えられたとする．ここで，ϕ_0 は $\phi(s_0)$ を指す．特徴ベクトル ϕ_0 を構成する要素 λ を出力 $Q_a(\phi_0)$ への影響の大きさに基づき識別する問題を考える．簡単のために出力クラスの出力が線形関数の場合を考える．式 (1) に ϕ を入力とし Q_a を出力する線形モデルを示す．

$$Q_a(\phi) = w_a^{\mathrm{T}}\phi + b_a \tag{1}$$

ここで，w_a，b_a は線形モデルの重みベクトルとバイアスを指す．線形モデルの場合，w は ϕ の出力 Q_a への影響の大き

さを表す．

本研究で扱う CNN は ϕ を入力とし Q_a を出力する非線形モデルであり，線形モデルにおける考え方をそのまま転用することはできない．しかし，任意の ϕ_0 を入力とする非線形モデル $Q_a(\phi)$ の ϕ_0 近傍での近似は可能である．式 (2) に ϕ_0 近傍での $Q_a(\phi)$ の 1 次のテイラー展開を示す．

$$Q_a(\phi) \approx w^{\mathrm{T}}\phi + b \tag{2}$$

ここで，w は，ϕ_0 近傍での Q_a の ϕ による微分である式 (3) に等しい．

$$w = \left.\frac{\partial Q_a}{\partial \phi}\right|_{\phi_0} \tag{3}$$

式 (3) の w は各ピクセルの，CNN の出力 Q_a に対する勾配を示しており，w の各要素の絶対値が大きさは，それに対応するピクセルの出力への影響が大きさを表す．

m 行，n 列，チャネル数 c の画像 s が与えられた時の Saliency map $M \in \mathcal{R}^{m\times n\times c}$ の計算手順について説明する．はじめに，誤差逆伝搬法 (back propagation) で式 (3) を計算し w を求める．次に，求めた 1 次元ベクトル w の要素数は $m\times n\times c$ に等しい．w を $m\times n\times c$ の形に再配置することによって Saliency map を得る．Saliency map を可視化する際には，勾配の大きさを正規化し画像として出力する．このとき，勾配が大きい画像のピクセルは白色に近づく．

■ Guided Saliency map

Guided Saliency map は Saliency map と同じく，入力画像の各ピクセルが CNN の出力に及ぼす影響の大きさを可視化する手法である．Saliency map は誤差逆伝搬法で勾配を求めるが，Guided Saliency map は guided backpropagation とよばれる制約付きの誤差逆伝搬法で勾配を求める．

Guided Saliency map は，高層の特徴マップから CNN のノードの出力を guided backpropagation で逆伝搬させ，画像を再構築する．逆伝搬によって再構築された画像は高層の特徴マップのノードの活性への影響を与える部分を示す．Figure 2 に逆伝搬による画像の再構築の概略を示す．

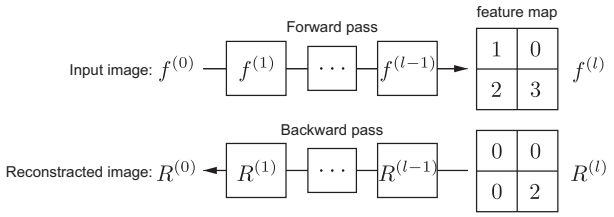


Figure 2: Image reconstraction

ここで，$f^{(l)}$, $R^{(l)}$ はそれぞれ CNN の l 層目の出力と逆伝搬を指す．guided backpropagation は誤差逆伝搬法に deconvolution と同様の制約を加わえた計算手法である．活性化関数に Relu を用いた場合の activation, backpropagation, deconvolution, guided backpropagation の計算式を以下に示す．

activation:

$$f_i^{(l+1)} = Relu(f_i^{(l)}) = \max(f_i^{(l)}, 0) \tag{4}$$

backpropagation:

$$R_i^{(l)} = (f_i^{(l)} > 0)\cdot R_i^{(l+1)}, \text{where } R_i^{(l+1)} = \frac{\partial f^{(L)}}{\partial f_i^{(l+1)}} \tag{5}$$

deconvolution:

$$R_i^{(l)} = (R_i^{(l+1)} > 0)\cdot R_i^{(l+1)} \tag{6}$$

guided backpropagation:

$$R_i^{(l)} = (f_i^{(l)} > 0)\cdot (R_i^{(l+1)} > 0)\cdot R_i^{(l+1)} \tag{7}$$

Figure 3 に活性化関数 Relu を適用した際の guided backpropagation の概略を示す．

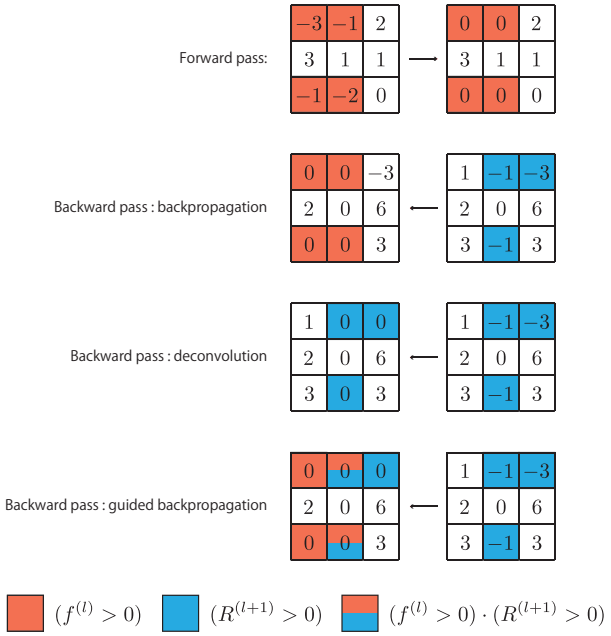


Figure 3: Guided backpropagation

Figure 3 下部の凡例はそれぞれの色のマスが受ける制約を示す．Figure 3 の赤マスは Relu，または Relu と同様の制約を受け，青マスは deconvolution の制約を受ける．赤と青の 2 色からなるマスは Relu と deconvolution 両方の制約を受ける．guided backpropagation で求めた勾配を Saliency map と同様の手法で可視化したものを Guided Saliency map とよぶ．

■ Grad-CAM

Grad-CAM は，畳み込み層または max-pooling 層の出力が CNN の出力に与える影響を可視化する手法である．いま，マップを求めたい畳み込み層が幅 u，高さ v の特徴マップを K 枚出力しているとする．求めたいマップ Grad-CAM は $L^a_{\mathrm{Grad-CAM}} \in \mathbb{R}^{u\times v}$ と表される．Grad-CAM を求めるために，特徴マップ F に関する出力 Q_a の微分 $\frac{\partial Q_a}{\partial F^k_{ij}}$ を求める．ここで，F^k_{ij} は k 枚目の特徴マップ F^k の i 行 j 列の要素を指す．求めた微分 $\frac{\partial Q_a}{\partial F^k_{ij}}$ の大域平均プーリングを求め重み α^a_k を得る．式 (8) に α^a_k の計算式を示す．

$$\alpha^a_k = \frac{1}{Z}\sum_i\sum_j \frac{\partial Q_a}{\partial F^a_{ij}} \tag{8}$$

ここで，Z は正規化項を指し，α^a_k は特徴マップ F^k の出力 Q_a への影響の大きさを示す．

Grad-CAM のヒートマップを，重み α^a_k で乗じた特徴マップ F^k を足しあわせることで求める．式 (9) にヒートマップ $L^a_{\mathrm{Grad-CAM}}$ の計算式を示す．

$$L^a_{\mathrm{Grad-CAM}} = Relu\left(\sum_k \alpha^a_k F^k\right) \quad (9)$$

$L^a_{\mathrm{Grad-CAM}}$ の計算に Relu を用いなかった場合，他クラスの出力に関係するピクセルが強調されることがある．ここで，Relu は，出力 Q_a を大きくするピクセルを求める役割を持つ．$L^a_{\mathrm{Grad-CAM}}$ を可視化する際には，マップの値を正規化し出力する．

■ Guided Grad-CAM

Grad-CAM と Guided Saliency map の要素積を Guided Grad-CAM とよぶ．Saliency map, Guided Saliency map と同様に勾配が大きいピクセルほど白色に近づく．

3. 提案法

状態 s が画像データで表される場合に，アフォーダンス Q_a を誘発する特徴 λ を識別する方法を提案する．提案法はアフォーダンスの学習と CNN の解析の 2 段階からなる．提案法の手順を Algorithm 1 に示す．

Algorithm 1 Feature identification

Input: Dataset $\{(s_1, Q^1_a), \ldots, (s_{\mathrm{N}}, Q^{\mathrm{N}}_a)\}$
Create a CNN model:
  Input: $\phi(s)$.
  Output: $(Q_a, 1 - Q_a)$.
Learn Affordance: Train CNN with Dataset.
Analyze CNN:
  Input: s.
  Select target output: Q_a or $1 - Q_a$.
  Apply: Saliency map, Guided Saliency map, Grad-CAM, Guided Grad-CAM.
Output: Discriminative features $\{\lambda_1, \ldots, \lambda_n\}$ in $\phi(s)$ for Q_a.

以下でアフォーダンスの学習と CNN の解析について説明する．

3.1 アフォーダンスの学習

設計物の画像データ群を入力，画像に対する利用者の知覚のデータ群を教師データとして CNN にアフォーダンスを学習させる．教師データは 0, 1 の 2 値とし，出力層の活性化関数は，出力層の全ノードの出力値の合計が 1 となる softmax 関数とする．式 (10) に softmax 関数の定義式を示す．

$$y_i = \frac{e^{x_i}}{\sum_{n=1}^{\mathrm{N}} e^{x_n}} \quad (10)$$

ここで，x_i, y_i はそれぞれ i 番目の出力ノードの活性化関数適用前，適用後の値を示す．

3.2 CNN の解析

アフォーダンスを学習した CNN を解析し, アフォーダンスを誘発する特徴 λ を識別する．解析対象とする出力を選択し，2.2 節で述べた Saliency map, Guided Saliency map, Grad-CAM, Guided Grad-CAM の四つの手法を適用する．

4. 実験

4.1 実験設定

本実験では，被験者が画像の座れるアフォーダンスの有無を判別し, アフォーダンスの学習に用いる教師データを作成した．画像データには画像認識のベンチマークとして用いられるデータセット The CIFAR100[Krzhevsky 09] の画像を用いた．実験に際し，CNN の学習に用いる訓練データ (train data) と，学習後の CNN の評価に用いるテストデータ (test data) を作成した．Table 1 に作成したデータの内訳を示す．

Table 1: Data summary

Data	Sittable:True	Sittable:False	Sum
Train	1824	48176	50000
Test	361	9639	10000

Table 1 のセルは画像データは座れるアフォーダンスを備える画像，備えない画像の枚数を示す．訓練データとテストデータのデータ数はそれぞれ 5 万，1 万個である．実験に用いた CNN の構造の詳細を以下に示す．

Table 2: Architecture of the CNN learning affordance

Layer name	Layer description
input	Input 32 x 32 RGB image
conv1	3 x 3 conv. 32 ELU, stride 1
conv2	3 x 3 conv. 32 ELU, stride 1
max-pooling1	max-pooling stride 2
conv3	3 x 3 conv. 64 ELU, stride 1
conv4	3 x 3 conv. 64 ELU, stride 1
max-pooling2	max-pooling stride 2
dense	dense. 512 ELU, dropout0.5
softmax	2-way softmax

活性化関数，最適化関数はそれぞれ ELU(Exponential Linear Units), Adam とし，誤差関数には交差エントロピー (cross entropy) を用いた．

4.2 実験結果

Table 3 に訓練データとテストデータに対するアフォーダンス学習後の CNN の予測の精度，再現率，F 値を示す．

Table 3: Evaluation of CNN's predicts

Data	Precision	Recall	F-measure
Train	0.922	0.994	0.956
Test	0.826	0.776	0.800

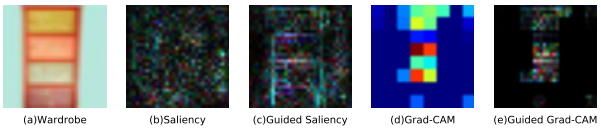
Table 3 の結果から，CNN は座れるアフォーダンスを学習しているといえる．

Figure 4 に Saliency map, Guided Saliency map, Grad-CAM, Guided Grad-CAM による CNN の解析結果を示す．

Figure 4 は，解析対象とする出力を $\max(Q_a, 1 - Q_a)$ として解析した結果である．Figure 4 のサブキャプション，Ground truth, Predict はそれぞれ入力画像 (a) の座れるアフォーダンスの有無の CNN の予測結果を示す．True は画像が座れるアフォーダンスを備えることを，False は画像が座れるアフォーダンスを備えないことを意味する．Figure 4 の (b)∼(e) はそれぞれ CNN の解析法を示す．

Grad-CAM, Guided Grad-CAM は任意の畳み込み層，max-pooling 層に対して適用可能な解析法である．Figure 4 に示す Grad-CAM, Guided Grad-CAM の結果は，Table 2 に示す max-pooling2 に対する解析結果である．

(1)Ground truth:True Prdict:True



(2)Ground truth:False Predict:True

(3)Ground truth:True Predict:False

Figure 4: Identification of features inducing affordance

5. 考察

Figure 4 の結果を画像中の設計物の部位を対応させ考察する．Table 4 に設計物の部位に対応させた結果を示す．

Table 4: Equivalent parts for features inducing affordance

method	Equivalent parts		
	(1)	(2)	(3)
Saliency	-	-	-
G.Saliency	plane	side, horizonal	leg
Grad-CAM	leg, side	-	-
G.Grad-CAM	leg, side	-	leg

ここで，Table 4 の番号 (1)〜(3) は Figure 4 の番号に対応し，Saliency, G.Saliency, G.Grad-CAM はそれぞれ Saliency map, Guided Saliency map, Guided Grad-CAM を指す．Figure 4-1 は設計物の脚部，側部，平面部が座れるアフォーダンスを誘発することを，Figure 4-2 は設計物の側部，水平部が座れるアフォーダンスを誘発することを示す．Figure 4-3 は設計物の脚部を認識していることを示すが，CNN は Figure 4-3 の (a) が座れるアフォーダンスを備えないという誤った予測をしている．

Figure 4-1 と Figure 4-3 の設計物は脚部の形状が異なる．また，Figure 4-2 では棒状の側部が座れるアフォーダンスを誘発しており，座れるアフォーダンスを誘発する特徴の 1 つに垂直な棒状の脚部があるといえる．

Table 4 に示すように Figure 4-1 では Grad-CAM と Guided Grad-CAM が, Figure 4-2 では Guided Saliency map の結果が最も多くの部位に対応可能であり，特徴識別に最も優れた解析法の特定は困難である．しかし，複数の解析法を統合的に利用する方法が考えられる．

Figure 4-2 の (c) と (e) では特徴識別の結果が大きく異なる．これは畳み込み層と max-pooling 層での演算によって生じる．Figure 5 に Figure 4-2 の (a) を入力し，CNN を構成する各畳み込み層，max-pooling 層に Grad-CAM を適用した結果を示す．

Figure 5 の結果は，畳み込み層と max-pooling 層を伝搬する過程で勾配が大きい特徴の位置が変化することを示す．Guided

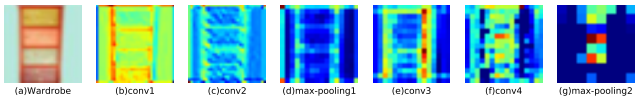


Figure 5: Grad-CAM results for each layer of CNN

Saliency map と Grad-CAM, Guided Grad-CAM による特徴識別の結果が大きく異なる場合は，Grad-CAM, Guided Grad-CAM を適用する層を適宜変更し，再度 Guided Saliency map と比較し，考察する必要がある．

6. まとめと今後の課題

本研究では，これまで概念定義にとどまっていたアフォーダンスを定式化し，アフォーダンスを誘発する特徴を識別する方法を提案した．具体的には，設計物の画像データを入力，利用者が知覚した座れるアフォーダンスを教師データとして学習した CNN を Saliency map, Guided Saliency map, Grad-CAM, Guided Saliency map で解析した．座れるアフォーダンスを誘発する特徴を識別した結果，設計物の「脚部」が座れるアフォーダンスを誘発することがわかった．

特徴識別の結果は手法によって異なり，特徴識別に最も優れた手法の特定は困難である．しかし，複数の特徴識別手法を統合的に利用する方法も考えられる．Grad-CAM, Guided Grad-CAM の結果が，解析対象とする中間層の選択に大きく依存するサンプルが複数認められた．そのため，これら二つの解析法を特徴識別に用いる際には，CNN の構造や入力画像のピクセル数を考慮しなければならない．今後の課題として，時系列データや 3 次元モデルデータを対象とした特徴識別がある．

参考文献

[Norman 99] Norman, Donald A. “Affordance, conventions, and design.” interactions 6.3 (1999): 38-43.

[LeCun 89] LeCun, Yann, et al.“Backpropagation applied to handwritten zip code recognition.” Neural computation 1.4 (1989): 541-551.

[Simonyan 13] Simonyan, Karen, Andrea Vedaldi, and Andrew Zisserman. “Deep inside convolutional networks: Visualising image classification models and saliency maps.” arXiv preprint arXiv:1312.6034 (2013).

[Springenberg 14] Springenberg, Jost Tobias, et al. “Striving for simplicity: The all convolutional net.” arXiv preprint arXiv:1412.6806 (2014).

[Selvaraju 16] Selvaraju, Ramprasaath R., et al. “Grad-CAM: Why did you say that? Visual Explanations from Deep Networks via Gradient-based Localization.” arXiv preprint arXiv:1610.02391 (2016).

[Krzhevsky 09] Krizhevsky, Alex, and Geoffrey Hinton. “Learning multiple layers of features from tiny images.” (2009).