2L3-OS-09b-1

DCNNを用いた画像の質感認知ー音象徴性からのアプローチー

Texture Recognition of Material Images using DCNN - Approach from Sound Symbolism

權 眞煥[*1] 川嶋 卓也[*1] 下田 和[*1] 坂本 真樹[*1]
Jinhwan Kwon Takuya Kawashima Wataru Shimoda Maki Sakamoto

[*1] 電気通信大学大学院 情報理工学研究科
Department of Informatics, The University of Electro-Communications, Tokyo, Japan

Material (or texture) recognition represents that various physical quantities belonging to objects are detected by human sensory receptors and processed in the brain. However, it is difficult to define material recognition because the amount of information is enormous in material and texture. In recent years, some researchers have raised a possibility that perceptual characteristics of textures can be integrally expressed by sound symbolism. Sound symbolism represents a phenomenon in which a certain amount of information perceived from the physical world is strongly associated with phonological elements in the brain. In this research, we aim to generate material and texture expressions using sound symbolic words as variables to converge various material and texture features and Deep Convolutional Neural Network (DCNN). As a result, it became possible to output stochastic phoneme of sound symbolism to the input image. Our achievement is expected as a new method capable of expressing various fine texture information comprehensively.

1. はじめに

世の中に存在しているモノはその特有の質感を持つ. 質感認知とは, モノに属している様々な物理量(表面形状, 形, 色など)が人の知覚センサーに検知され, 脳で処理される一連の流れから成り立つ[小松 12]. しかし, 質感はその情報量が膨大であるため, 質感認知を定義することは困難であった[小松 12]. それゆえ, 質感は光沢感, 透明感, 乾湿感, 粗滑感などに細かく分類され, 様々な分野で研究がされてきた. ここで, 近年清水らは質感の統合表現として音象徴性を利用し, その妥当性や有効性を検証した[清水 14]. 音象徴性とは外から知覚されるある情報量が脳の中の音の要素(音韻)と強く結びつく現象を表す. つまり, 人は様々な知覚量をある音韻に収束させ, 学習を行っていることを示す. この学習過程では偏向的な質感認知ではなく, 微細かつ様々な質感情報を統合的に表現していることとつながる. 本研究では, この学習過程と知覚特性を用いて, 質感情報を引き出す新たな手法を試みる.

ニューラルネットワークとは生物の脳内の神経回路の仕組みを模したモデルであり, 深層畳み込みニューラルネットワーク(Deep Convolutional Neural Network: DCNN)は一般物体認識の分野で著しく発展を遂げ, 学習の過程で自ら有効な特徴を抽出すると報告されている[岡谷 15]. しかし, 従来の研究ではCNN を使用して素材分類などの研究が進められているが, 画像がもつ質感そのものの表現については至っていない. 特に, 質感は光沢感, 透明感, 乾湿感, 粗滑感などの様々な特徴を含んでいるため, この様々な質感特徴を収束させる変数が必要であり, 質感表現を適切に扱える工学的な手法が求められる. 本研究では, 様々な質感特徴を収束させる変数としての音象徴語と, DCNN を使用し, 質感表現を生成することを目的とする.

連絡先:坂本真樹, 電気通信大学大学院情報理工学研究科, 〒182-8585 東京都調布市調布ヶ丘 1-5-1 電気通信大学西 6 号館 511 室, Tel/Fax: 042-443-5535, maki.sakamoto@uec.ac.jp

2. 関連研究

2.1 質感と DCNN に関する研究

CNN は従来の機械学習とは異なり, 有効な画像特徴量を学習の過程で自動で抽出することができ, 2012 年の The ImageNet Large Scale Visual Recognition Challenge 以降, 一般物体認識分野において大きな注目を浴びている. Cimpoi らは, 一般物体認識の 1,000 種類のデータが学習された CNN の活性化信号を画像特徴量として使用する Deep Convolutional network Activation Features と Improved Fisher Vector を用いて, FMD 画像の素材分類を行い, 認識率を 67.1%に向上させ, 質感認知と CNN の可能性を示した[Cimpoi 14]. 一方, 2014 ImageNet Large Scale Visual Recognition Challenge(ILSBRS)では Google が誤差率 6.6%, 2015 ILSBRS では MSRA が誤差率 3.5%という結果を出すなど, 物体カテゴリ認識や画像認識の分野において大きな成果を出しており, ドロップアウトなど学習の際に有効な手法が報告されている[Srivastava 14].

2.2 オノマトペに関する研究

清水らは曖昧とされてきたオノマトペの微細な印象を, 客観的に推定する手法を提案した[清水 14]. オノマトペの特徴である音象徴性に着目し, 被験者実験によって得た印象評価値から数量化理論 I 類によってカテゴリ数量を決定することによって, 高い予測精度を有するオノマトペ表現の印象評価手法を実現することができた. さらに, 同システムを利用して, 質感に関連する触覚のカテゴリと音象徴語の関係を可視化する研究も進められている. 特に, 素材―知覚カテゴリ―音象徴語の関係性を可視化することにより, 音象徴性の直感的な活用の有用性が検証されつつある.

3. 実験設計

3.1 データセット

本研究では,『Flickr Materials Database (https://people.csail.mit.edu/celiu/CVPR2010/FMD/)』1000 枚から切り出された 1946

枚(図 1 参照)と1枚あたり 10 人の被験者より複数回答されたオノマトペ 30138 語のデータから，DCNN の学習に適したオノマトペ 29443 語を用いた．

図 1. 切り出された FMD 画像例

3.2 学習手法

本研究では，畳み込みニューラルネットワークのモデル構成として VGG16 を参考とした[Simonyan 14]．また fc6,fc7 は Drop Out を行うことで過学習を考慮し，fc8 の出力関数については 0 から 1 に収まるように sigmoid 関数を採用した(図 2 参照)．誤差関数は二乗誤差関数，勾配降下法として確率的勾配降下法(Stochastic Gradient Descent, SGD)，バッチサイズ 32 のミニバッチ学習を行った．このときパラメータは学習係数=0.001，epoch=1000 とし，2015 年のオックスフォード大学が行った 1000 種類の一般物体認識の 16 の学習モデルを用いて，ファインチューニングを行った[Simonyan 14]．

学習においては，1枚の画像に対して複数の正解データによる学習を行った．その正解データはオノマトペを表現する 73 次元配列(「反復あり/なし(番号 0～1)」「1モーラ目の母音(2～6)」「1モーラ目の子音(7～33)」「1モーラ目の特殊語尾(34～36)」「2モーラ目の母音(37～41)」「2モーラ目の子音(42～68)」「2モーラ目の特殊語尾(69～72)」)に沿った形式に分解し，1枚の画像に対してオノマトペ1語ずつ対応付けた．これは画像に対して受ける質感の多様性を考慮するためである．

図 2. 畳み込みニューラルネットワークの層構造

4. 結果

本研究では質感特徴を収束させる変数として音象徴性と，DCNN を使用し，質感表現を生成する試みを行った．画像を入力して，音象徴性を出力するために回答されたオノマトペを，73 次元の要素に分解し，それらを正解データとして学習を行った．その結果，学習係数 0.001，epoch 数 1000 の学習を行ったところ，図 3 となり， loss が減少していることから学習が確認され，質感に関する画像と音韻に関連性があることがわかった．

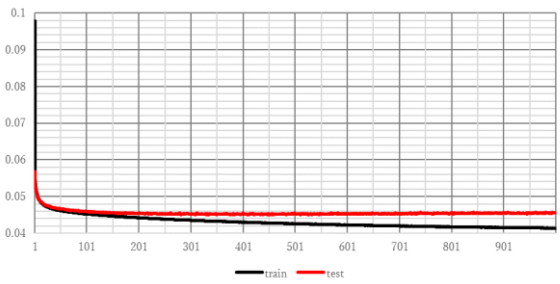

図 3. 学習曲線 (黒線が学習データ，赤線がテストデータ)

表 1 は図 1 の左の画像を学習した項目毎の出力値であり，入力画像に対してオノマトペを構成する 73 次元配列(音韻と構造)を確率的に出力することができた．また項目ごとの最も高い割合の要素を選択してオノマトペを作成すると「ごわごわ」となり，次の割合を組み合わせると「ざらざら」「もさもさ」「ぼこぼこ」なども抽出される結果となった．今回可能となった確率的な音韻出力は質感の多様性が保たれる新たな表現手法とつながる．

表 1. 図 1 の左画像を学習した項目毎の出力値

反復		
1	Re	0.96
2	noRe	0.04

1モーラ目								
母音			子音			特殊語尾		
1	o	0.61	1	g	0.34	N	0.02	
2	a	0.31	2	z	0.22	Q	0.01	
3	u	0.09	3	m	0.15	R	0.01	
4	i	0.05	4	b	0.10			
5	e	0.02	5	h	0.09			
			6	p	0.08			
			7	k	0.03			
			8	s	0.03			
			9	t	0.01			
			10	w	0.01			

2モーラ目								
母音			子音			特殊語尾		
1	a	0.82	1	w	0.38	N	0.01	
2	o	0.12	2	r	0.19	Q	0.02	
3	u	0.08	3	s	0.13	R	0.01	
4	i	0.05	4	k	0.09	L	0.01	
5	e	0.02	5	t	0.08			
			6	ty	0.03			
			7	h	0.01			
			8	p	0.01			
			9	b	0.01			
			10	y	0.01			

5. 結論

本研究では，様々な質感特徴を収束させる変数としての音象徴語と，DCNN を使用し，質感表現を生成することを試みた．その結果，入力画像に対して音象徴性の確率的な音韻出力が可能となった．この成果は今後微細かつ様々な質感情報を統合的に表現できる新たな手法として期待される．

謝辞

本研究は JSPS 科研費 JP23125510, JP 25135713, JP15H05922 の助成を受けたものです．

参考文献

[小松 12] 小松 英彦: 質感の科学への展望，映像情報メディア学会誌, Vol. 66, No. 5, pp. 331-337 (2012).

[清水 14] 清水 祐一郎，土斐崎 龍一，坂本 真樹: オノマトペごとの微細な印象を推定するシステム，人工知能学会論文誌, Vol. 29, No. 1, pp. 41-52 (2014).

[岡谷 15] 岡谷 貴之: 機械学習プロフェッショナルシリーズ「深層学習」,講談社(2015).

[Cimpoi 14] Cimpoi, M., Maji, S., Kokkinos, I., Mohamed, S., & Vedaldi, A.: Describing textures in the wild. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3606-3613 (2014).

[Srivastava 14] Srivastava, N., Hinton, G. E., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 15(1), 1929-1958. (2014).

[Sharan 09] Sharan, L., Rosenholtz, R., Adelson, E.: Material perception: What can you see in a brief glance? Journal of Vision, Vol. 9, No. 8, pp. 784-784 (2009).

[Simonyan 14] Simonyan, K., & Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In Proc. International Conference on Learning Representations http://arxiv.org/abs/1409.1556, (2014).