

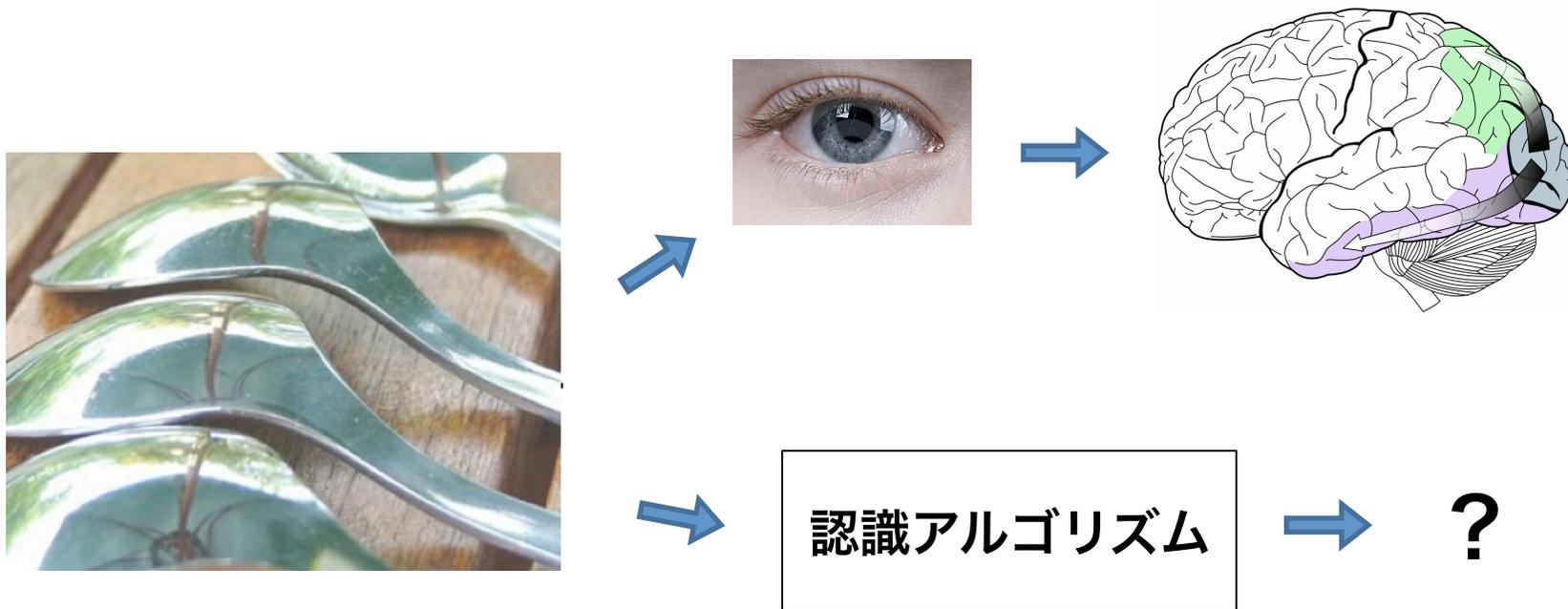
B01-1班

画像と言語を用いた質感情報表現の ディープラーニング

代表者：岡谷貴之，分担者：山口光太
東北大学

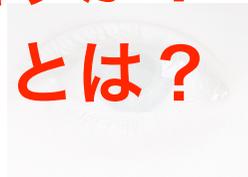
目標：質感認識システムを作る

- 質感を画像から認識する
 - 光沢度, 透明度, 柔らかさ, 滑らかさ, ...
 - 大量のデータを用いて学習



目標：質感認識システムを作る

- 質感を画像から認識する
 - 光沢度, 透明度, 柔らかさ, 滑らかさ, ...
- **2つの課題** 大量のデータを用いて学習
 - 画像から何を取り出すか? (特徴量)
 - 質感を「認識する」とは? (出力)



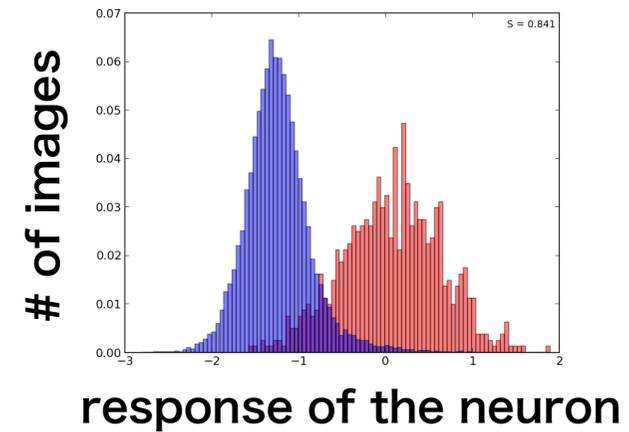
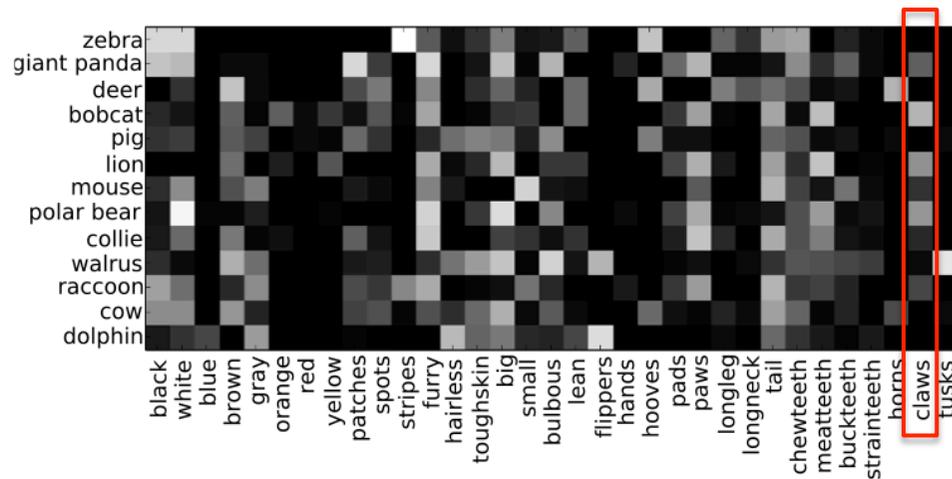
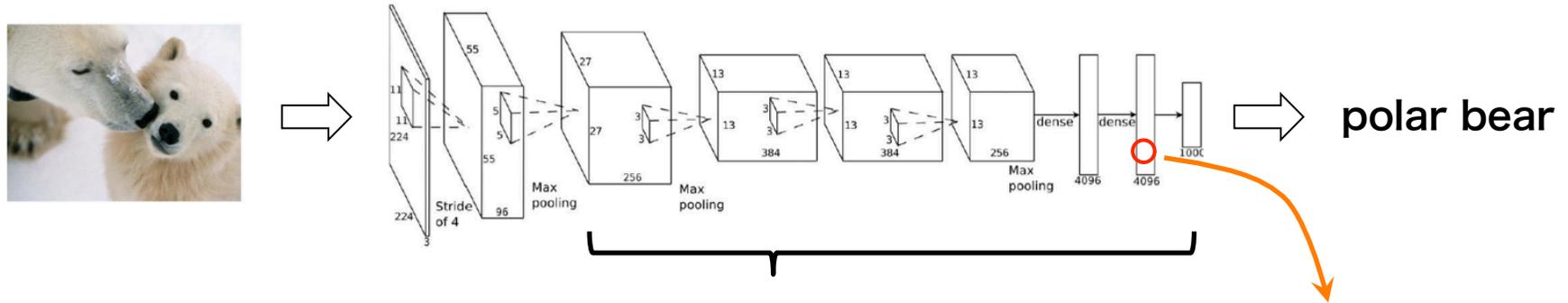
認識アルゴリズム



?

ディープニューラルネットの内部表現

Ozeki, Okatani, Understanding Convolutional Networks in Terms of Category-level Attributes, ACCV2014

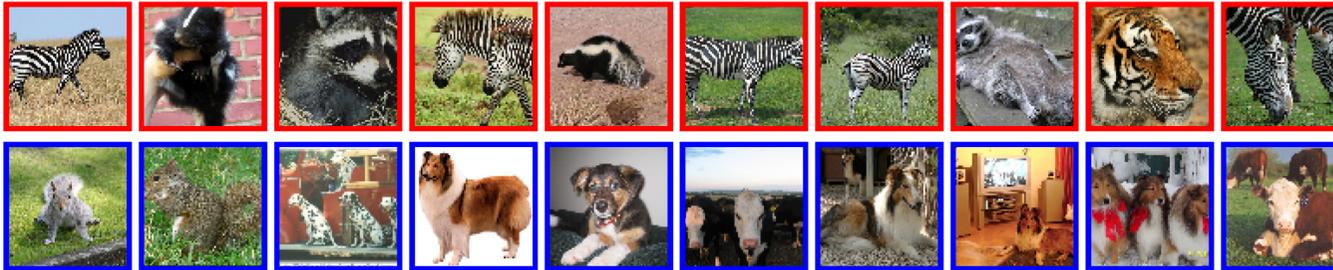


ディープニューラルネットの内部表現

Ozeki, Okatani, Understanding Convolutional Networks in Terms of Category-level Attributes, ACCV2014

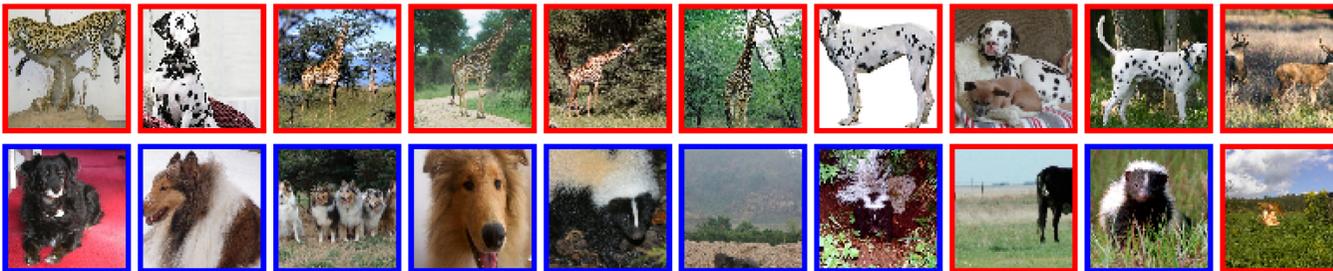
“stripes”

top
|
bottom



“spots”

top
|
bottom



“hairless”

top
|
bottom



材質の画像認識

[Liu+2010][Hu+2011][Timofte+2012][Sharan+2013]...



(a) Fabric



(b) Glass



(c) Leather



(d) Metal



(e) Paper



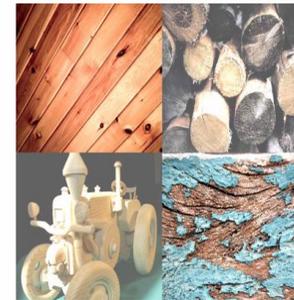
(f) Plastic



(g) Stone



(h) Water



(i) Wood



(j) Foliage

材質の画像認識

例：Fabric



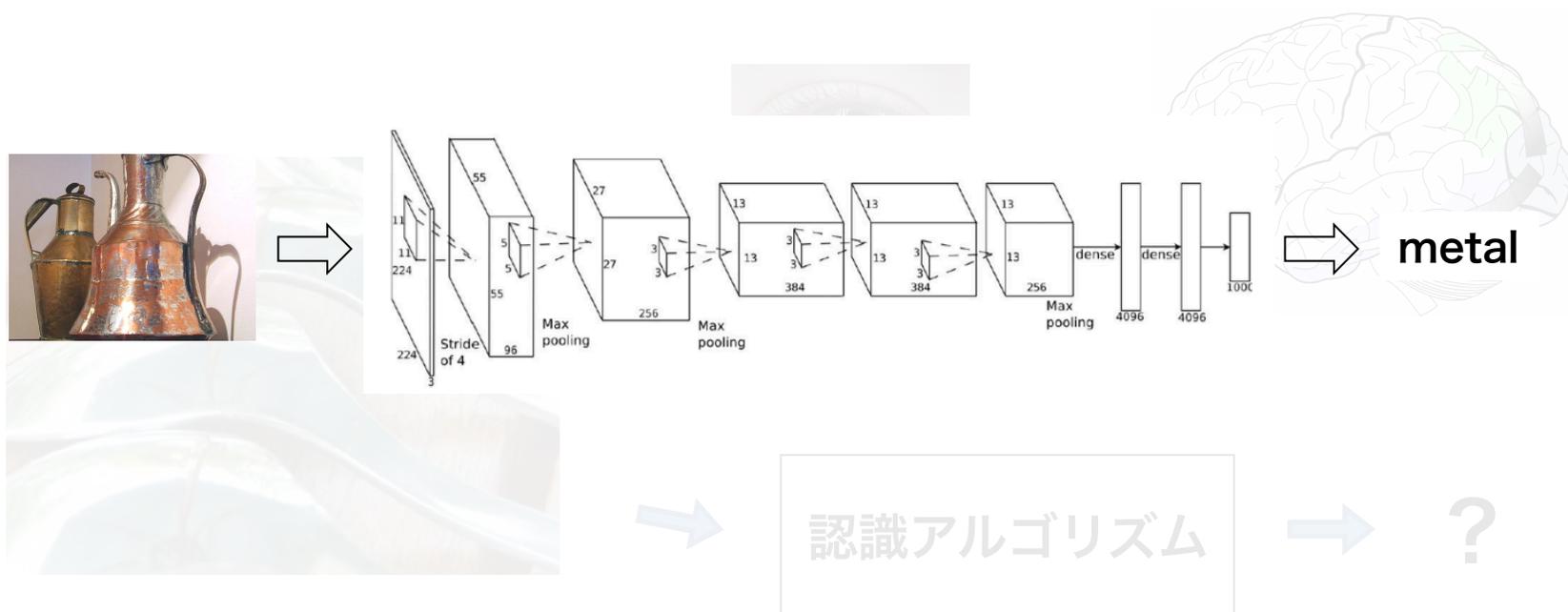
材質の画像認識

Liu, Ozay, Okatani, preparation for submission

Publication	Low level features	High level feat	Classifier	Recognition accuracy
Liu+2010	SIFT, color, curvature, edge-ribbon etc.	BoF	LDA	44.60
Hu+2011	Kernel descriptors(five)	Efficient match kernel	SVM	54.00
Timofte+2012	BIFs(Basic Image Features)	SPM-like + Comparative reasoning	CRC (Collaborative Representation Classifier)	55.78
Sharan+2013	SIFT, color, curvature, edge-robbon etc.	BoF	SVM	60.6
↑	Human (Mechanical Turk)	?	?	84.9
Cimpoi+2015	Deep CNN(VGG-19layers)	←	←	82.5±1.5
Ours (preparation for submission)	Ensemble of multiple Deep CNNs	←	Hierarchical sparse decision fusion	84.3±1.9

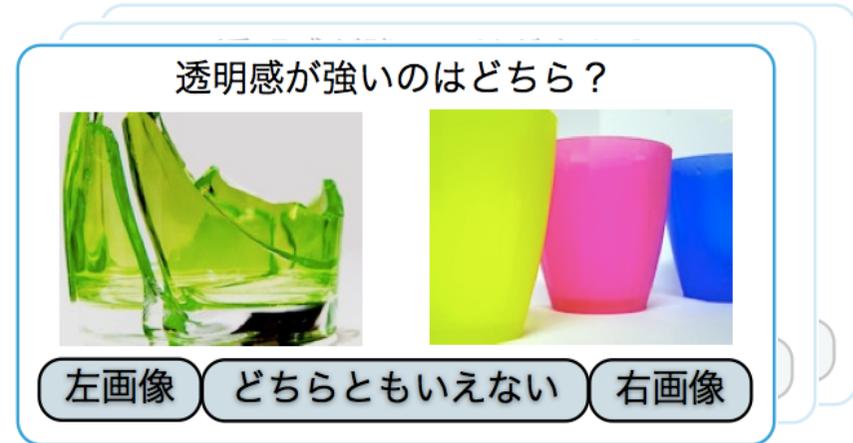
目標：質感認識システムを作る

- **2つの課題**
 - 画像から何を取り出すか？ (特徴量)...
 - 質感を「認識する」とは？ (出力)

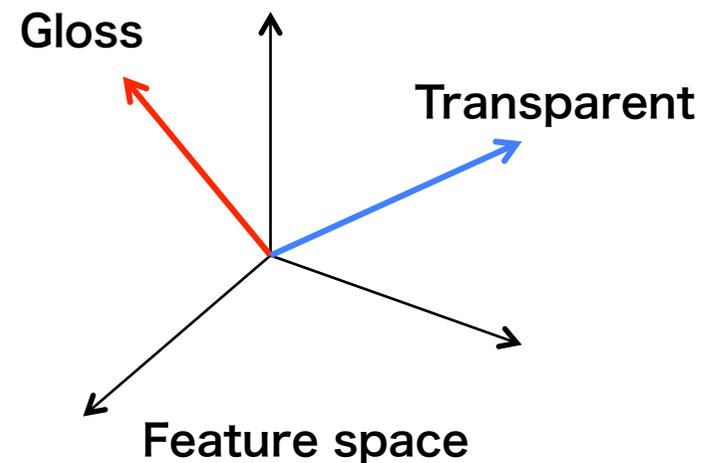
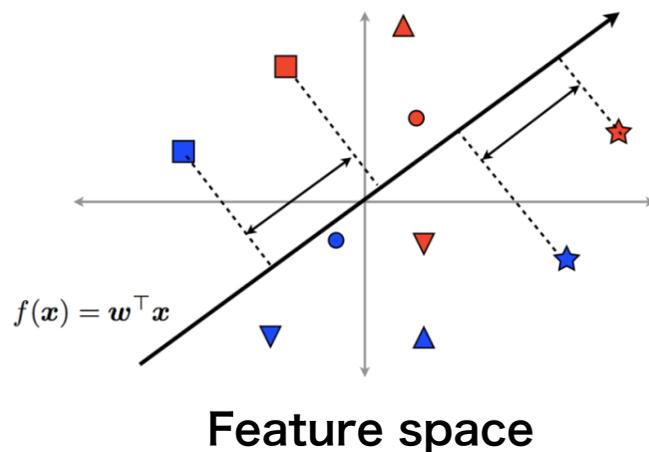


質感形容詞の予測

- 比較情報を人が作成
 - 13個の形容詞 (質感属性)
glossy, smooth,
transparent, cold,
fragile, hard, light,
wet, aged, beautiful,
clean, sticky, resilient



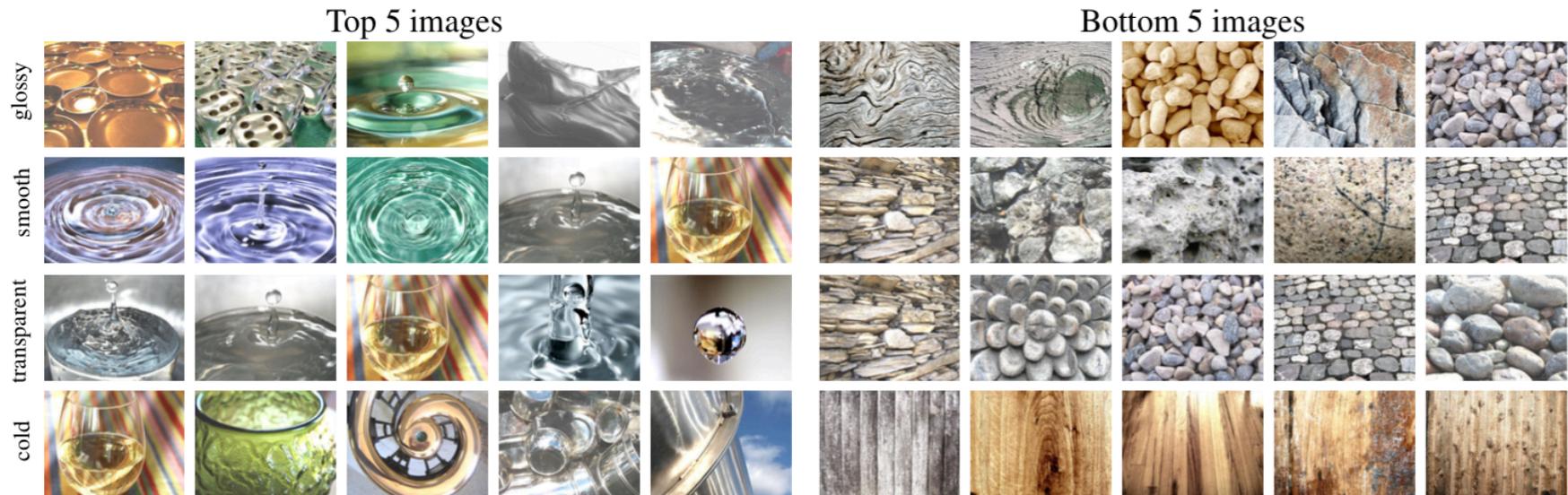
[Parikh-Grauman11]



質感形容詞の予測

	glossy	smooth	transparent	cold	fragile	hard	light	wet	aged	beautiful	clean	sticky	resilient
Pre-trained CNN	.857	.937	.924	.912	.911	.906	.901	.938	.929	.780	.905	.874	.948
Retrained CNN	.859	.948	.950	.902	.928	.906	.911	.936	.917	.800	.935	.934	.969
Fisher vector	.841	.937	.920	.907	.926	.908	.913	.944	.942	.803	.905	.894	.969
SIFT-BoF [1]	.780	.840	.770	.720	-	-	-	-	-	-	-	-	-
SIFT-BoF (dense) [1]	.770	.870	.780	.720	-	-	-	-	-	-	-	-	-
Neural Model [1]	.740	.750	.750	.650	-	-	-	-	-	-	-	-	-

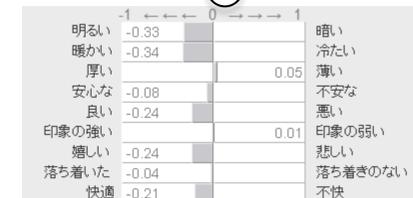
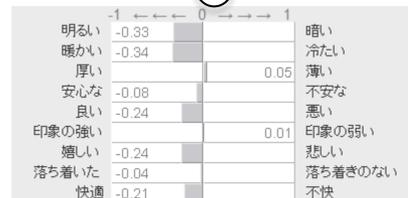
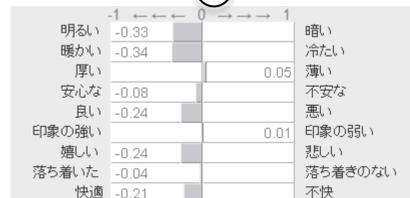
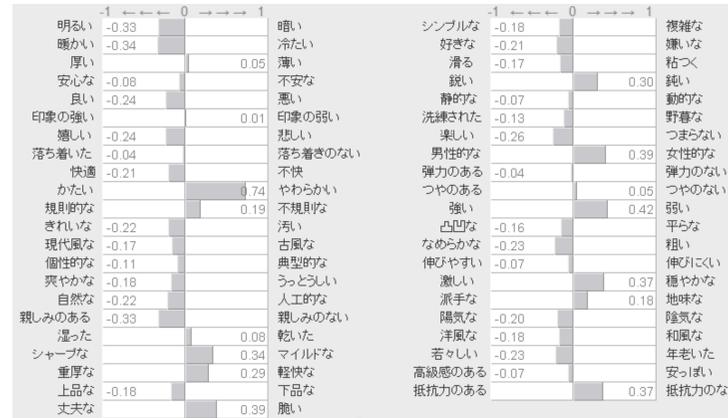
Table 2. Estimation accuracies of the pre-trained and retrained CNNs and Fisher vector for surface quality estimation. These three methods show almost the same performance, which is significantly better than the previous method (bag-of-features with SIFT+color statistics).



オノマトペの利用

土斐崎 龍一他, オノマトペと質感印象の結び付きに着目した商品検索への画像・テキスト情報活用の可能性, 2014

「ぴかぴか」 →



オノマトペの利用

土斐崎 龍一他, オノマトペと質感印象の結び付きに着目した商品検索への画像・テキスト情報活用の可能性, 2014

「ぴかぴか」

第1位:



第2位:



第3位:



第4位:



第5位:



「ごっごっ」

第1位:



第2位:



第3位:

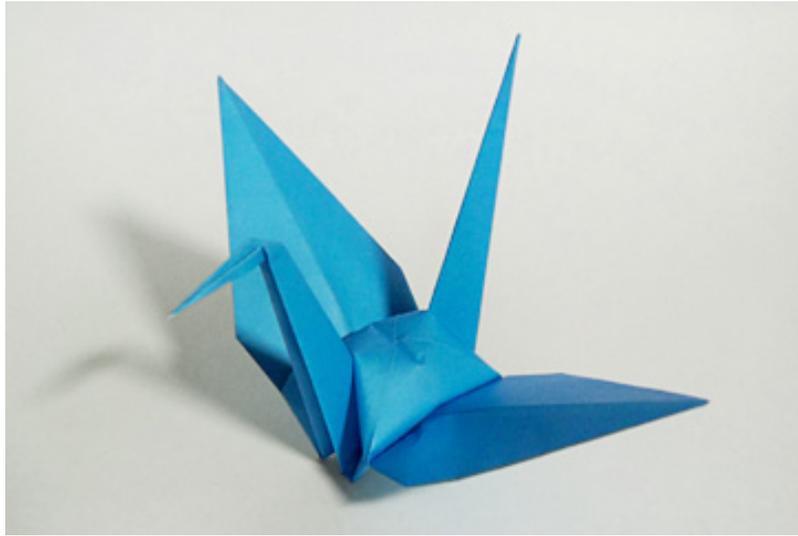


第4位:



第5位:

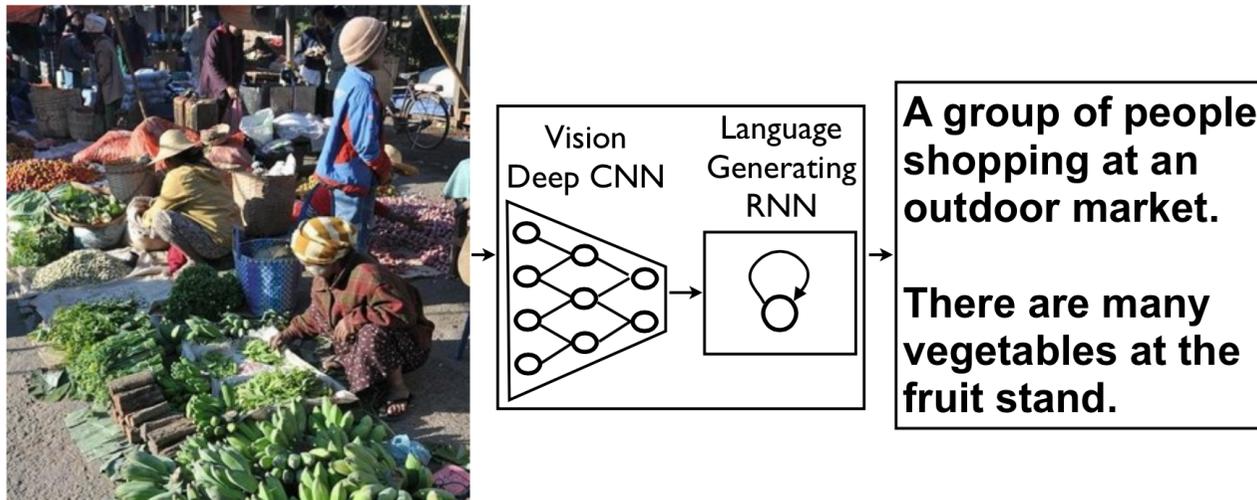




(ネットからの無断借用)

画像と言語の統合理解

- 目標：質感情報を含む言語記述の生成
- 画像記述 = 画像内の複数オブジェクト間の相互関係等の理解



Vinyals+, Show and Tell: A Neural Image Caption Generator, CVPR2015

まとめ

- **目標：質感を画像認識する工学的システムを実現**
 - 取り出すべき画像特徴は？
 - 質感を「認識する」とは？
- **ディープニューラルネットの利用**
 - 内部表現＝質感の表現空間
- **神経科学や心理物理との連携**