

# 深層学習による質感画像の認識・変換

第2回 質感のつどい

2016年11月30日

電気通信大学 大学院 情報学専攻 柳井 啓司 下田 和. 松尾 真

**国立大学法人電気通信大学** 

### 自己紹介: 柳井 啓司

- ・電気通信大学 大学院情報理工学研究科 情報学専攻 メディア情報学プログラム 教授
- ・研究分野:メディア情報学
  - ・画像・映像認識 特に、大規模Web画像・映像マイニング 最近は深層学習(ディープラーニング)を応用
  - · 質感研究歷: 2013~ (質感脳情報学2期目~多元質感計画班分担)

### [研究テーマ]

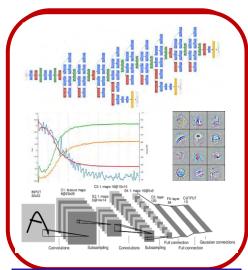
### 大量Web画像 と画像認識

- 1) 大量Web画像 を 大規模画像認識 に 利用.
  - 画像認識知識ベースの構築、深層学習に利用、
- 2) 画像認識技術 を 大量Web画像 に 適用.
  - Web画像検索, 自動分類, マイニング



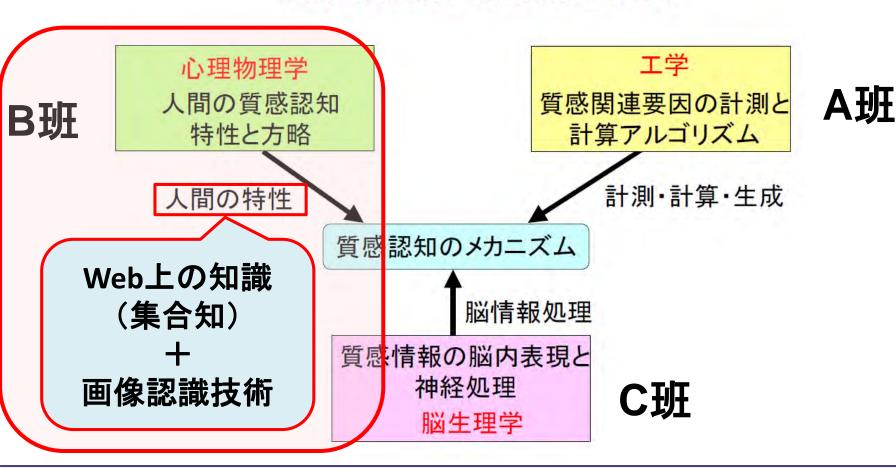






# 質感研究へのアプローチ: Web画像マイング

異分野連携による質感研究



# 研究例)画像領域エントロピー:単語概念の"視覚性"の指標 (2005)

- ・「動物」「哺乳類」vs.「ライオン」「パンダ」
- 「黄色い」「暗い」 vs. 「硬い」「騒がしい」 どちらが「視覚的」な言葉?どちらが画像から容易に認識可能?

視覚性(*visualness*)の指標の提案 *画像領域エントロピー* 

Keiji Yanai and Kobus Barnard: Image Region Entropy: A Measure of "Visualness" of Web Images Associated with One Concept, *Proc. of ACM International Conference on Multimedia 2005*, (2005/11).

### "視覚的"形容詞の調査結果

エントロピーの小さい単語

・"視覚性"が高い



7	purple	0.0412	
8	black	0.0443	
36	$\operatorname{red}$	0.9762	
39	blue	1.1289	
46	yellow	$1.2827$ $_{6}$ U	Eί

rankadjective.entropy1dark0.0118

2 senior 0.0118

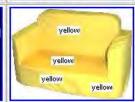
























-GT

electric

0.1411



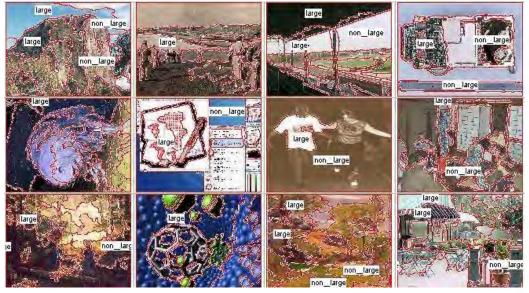


## 視覚性が高い: scary (怖い)

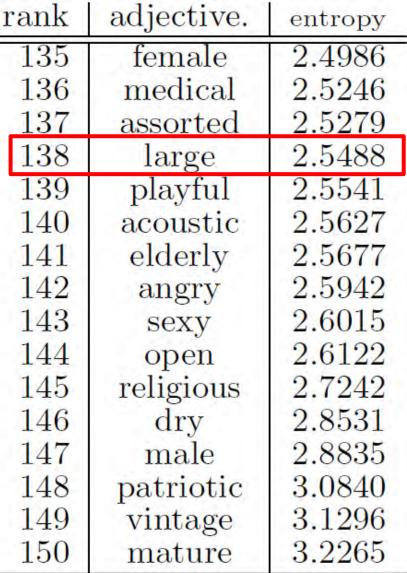


### "視覚的"形容詞の調査結果

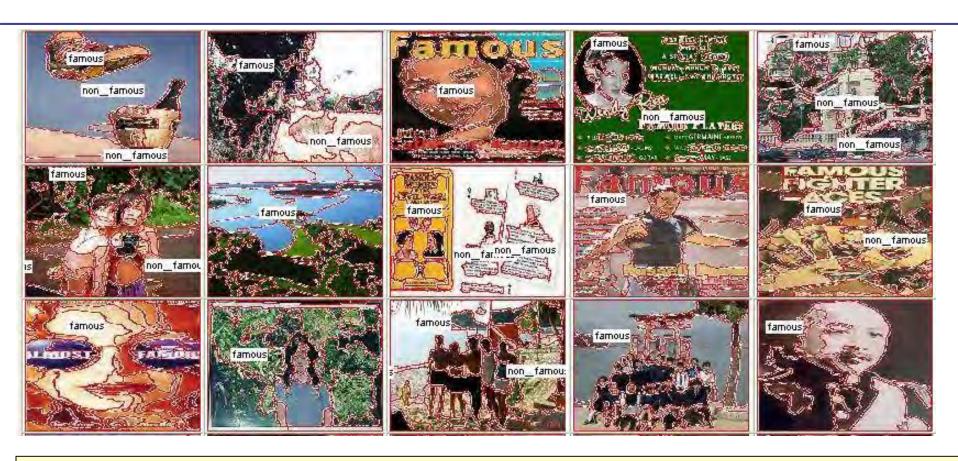
- エントロピーの大きい単語⇒ 分布にまとまっていない
- ・"視覚性"が低い



Large images



# 視覚性が低い:famous(有名な)



視覚性(visualness)の高い単語が画像認識に適している.

### 2013年の研究計画 (質感脳)

- ・質感表現に相当する言葉(主に形容詞)と 画像から抽出した特徴量の関係を Web上の「集合知」を用いて分析する
  - 本研究では、「集合知」=「タグ付き画像」

Web 上の大規模「集合知」データを活用した (新しい)「質感脳情報学」研究を実施

> 本アプローチの限界: 言語で表現できない質感は分析できない.



### 【2016年度研究計画】(多元質感知)

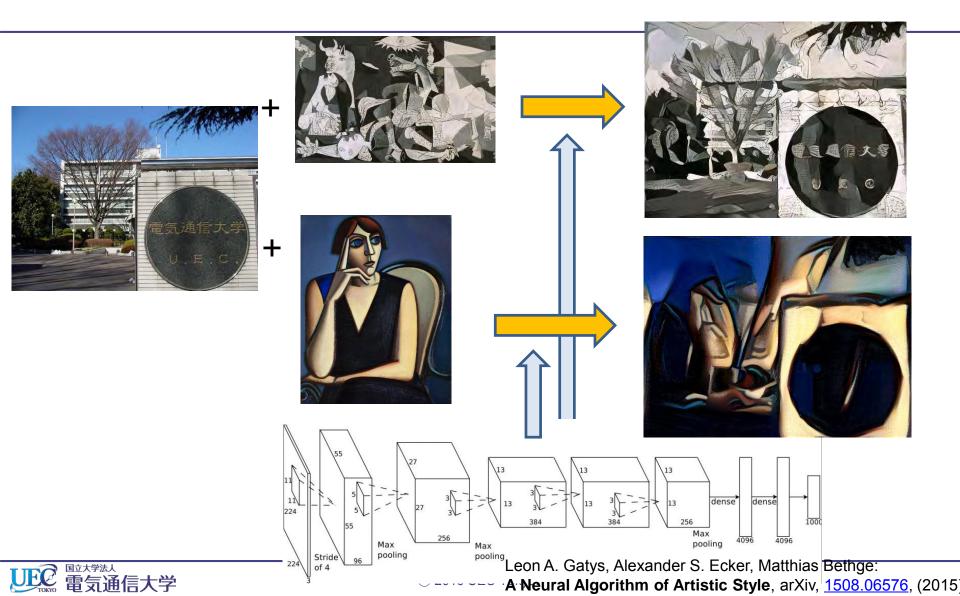
(西田班「信号変調に基づく視聴触覚の質感認識機構」分担者)

#### 「画像の質感変調」についても研究する

- (1) 深層学習によるスタイル変換技法を 用いた画像の質感操作
  - (A) 領域分割 + スタイル変換 による 任意部分の質感入れ替え
  - (B) Webからの自動スタイル収集と, 言葉による画像のスタイル変換の実現
  - (C) 高速マルチスタイル変換と, スタイル合成
- (2) 質感DBの作成 (主に画像のネット収集)



### 画像変換: スタイル転送





### 本日の内容

- ・Web画像を用いた質感画像認識と分析
  - 深層学習と質感認識
  - 「オノマトペ」に関する画像認識
  - 質感領域の領域分割: 弱教師あり
- ・画像スタイル変換による画像質感変調
  - スタイル変換.
  - 高速マルチスタイル変換.
  - -領域分割との組み合わせ.

# 深層学習による

# 質感画像認識



### 質感画像認識の研究

・テクスチャ認識: テクスチャクラスに分類





striped



素材認識:素材クラスに分類 e.g. FMD

fabric



stone



- ・ 形容詞 認識: 客観的(赤い) ⇔ 主観的 (美しい)
- ・属性認識: 角がある, 長い鼻, 長い耳, 尾っぽ

物体、シーンなどの名詞以外は、「質感」認識となる

## 質感画像認識も深層学習で

### 大幅性能向上

- ・物体と基本的に同じ
  - 2012まで、BOF+SVM
  - 2013から、CNN(畳み込みネット



- ・深層学習によって、認識性能が大幅向上
  - 例)素材認識:FMD (Flickr Material Database)
  - -2010 44.6% (color+SIFT)
  - -2014 65.6% (IFV+CNN)
  - 2016 **84.0%** (CNN with FT)

人間とほぼ同等の能力

Zhang, Y., Ozay, M., Liu, X., & Okatani, T. Integrating Deep Features for Material Recognition *ICPR* 2016.



### 人間を超えた画像認識

#### 1000種類 画像認識 コンペティション

・1000種類の物体のうちの1つが写った画像をコンピュータに見せて、名前を答えさせる課題。





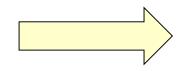


アリ (ant)





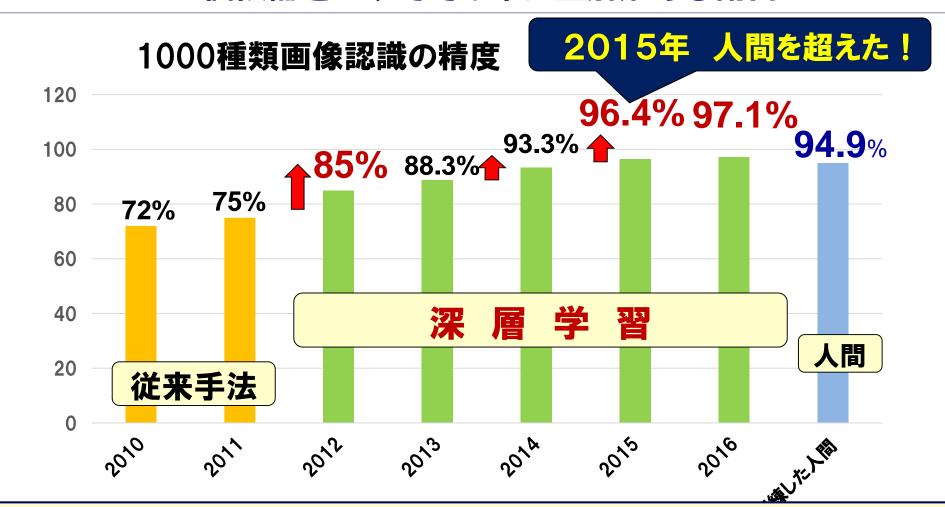
野球選手 (ballplayer)



障子 (shoji)

### 2010年以降の優勝チームの成績

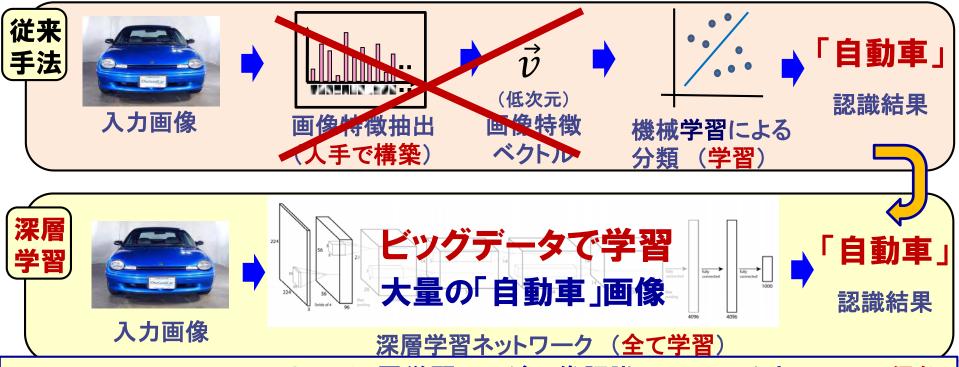
5個候補を上げてその中に正解がある割合



1000種類コンペティションは人工知能の進化を象徴している

### 深層学習が画像認識を変えた!

- ・2013年以降,画像認識の基本手法が 深層学習に置き換わった。
  - 画像認識研究者の夢であった「人間超え」があっさりと実現された.



Hand-crafted feature は悪、深層学習こそが画像認識の正しい手法⇒ Deep信仰



### あらゆるタスクに深層学習

・画像の認識のみでなく、応用面でも精度が飛躍

物体の認識

詳細認識







#### 画像検索



物体の位置の認識

#### 物体検出









#### 領域分割



デプス推定



キャプション生成



Q&A



(e) How many cats are here? 2

#### 画像合成



画像生成



# 深層学習と

# Web画像を用いた

質感画像分析

オノマトペ画像の分析

### 質感画像:オノマトペ画像

- 「オノマトペ」とは(B班 坂本さん)
  - 擬態語や擬音語.
    - ・しわしわ、ふわふわ、ざらざら など
  - 物体、素材、音などの状態、視覚のみならず



#### 五感すべての質感表現

素材カテゴリより, 多様な表現. 捉えどころがない. Challenging !!!

ふわふわ画像の例



### 報告内容

1. 質感画像データセット自動構築システム

- 2. Web収集画像を用いた質感語の視覚性評価実験
  - A)オノマトペに対応する画像の認識可能性評価
  - B) 122オノマトペ認識可能性ランキング.
  - C) 同一名詞内での、オノマトペの分類可能性の評価
- 3. CNN特徴(feature map)を用いた質感領域抽出

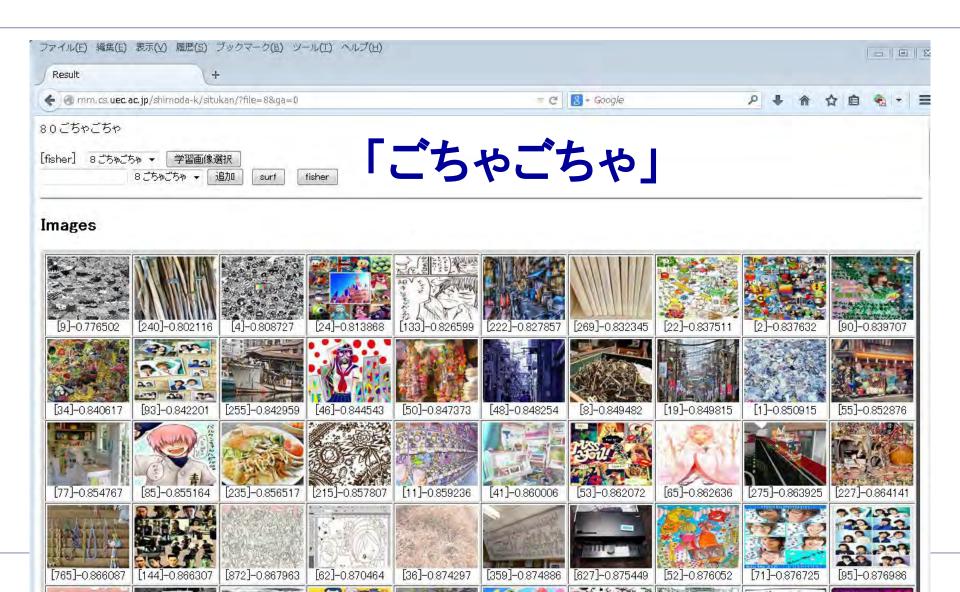


### 報告内容(1)

- ・ 質感画像データセット自動構築システム
- · Web収集画像を用いた質感語の視覚性評価 実験.
  - オノマトペに対応する画像の認識可能性評価
  - 122オノマトペ認識可能性ランキング.
  - 同一名詞内での、オノマトペの分類可能性の評価
- ・ CNN特徴 (feature map) を用いた質感領域抽出



### 質感画像データセット構築システム





#### データセット構築システム(yanai lab)



ホーム(検索結果) 自動データセット データセットの追加 データセットの削除 画像のアノテーション アノテーションの確認

Japanese English

もふもふ 5/31
query: もふもふ ▼ Co-occurrence: 0 ▼ cnn model: vgg16 exist ▼

States(overall) detail

Status(each) detail

Images page 0 1 2 3 4 5 6 7 8 9



## 質感画像データセット 構築システムのデモ

- http://mm.cs.uec.ac.jp/shimodak/shitsukan/
- ・特徴
  - Web画面でキーワードの入力するのみで画像収集.
  - 1000枚集めて、自動/手動ランキングし、 上位50/100枚をZIPファイルで一括ダウンロード可
  - Deep convolutional neural network activation feature (DCNN特徴) を用いた高精度リランキング.
  - 最大20台のPCによる並列処理で約5分で収集完了



### 自動画像収集システム

- ・約5分でデータセットを生成し、評価
  - Deep Learning特徴は計算コストがかかる. 1枚約3秒



- システムではクラスタマシンを20台使用し約1時間かかる処理を5分程度に短縮.



URL http://mm.cs.uec.ac.jp/shimoda-k/shitsukan/



### オノマトペ画像の収集

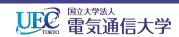
- bingAPIを利用
- 一つのクエリにつき1000枚画像を収集



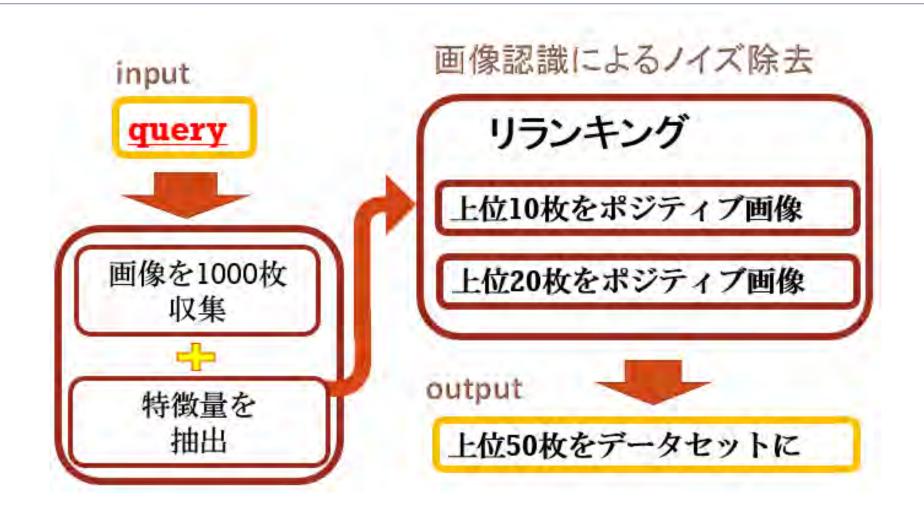
検索結果には上位 でもノイズ画像が含 まれてしまう (モクモク検索結果 上位50枚)

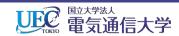


リランキング (re-ranking)



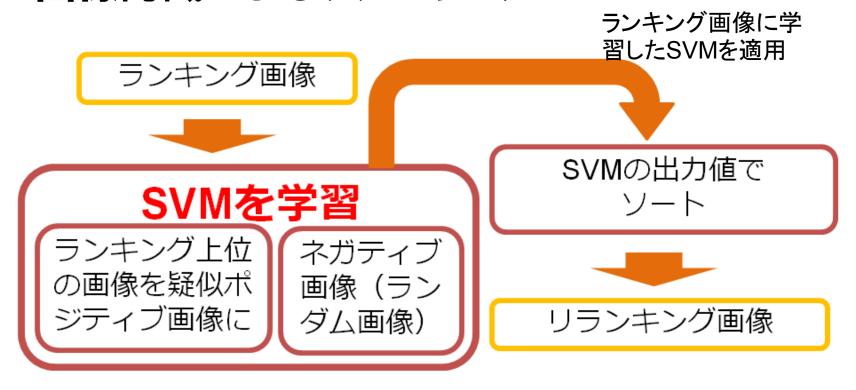
### データセットの構築





### リランキング手順

・画像認識によるリランキング



このリランキングを二回繰り返す



### リランキングの例

Queryを入力し、bingAPIでWeb画像を収集 「ざらざら」



Bing Image Search API 検索結果 上位50まで(ざらざら)



### リランキングの例

・検索結果の上位10枚を 使ってリランキング







リランキング結果その1(ざらざら)



### リランキングの例

### リランキング結果の 上位20枚でリランキング







リランキング結果その2 「ざらざら」画像セット完成!



### 実験(1):オノマトペ画像収集

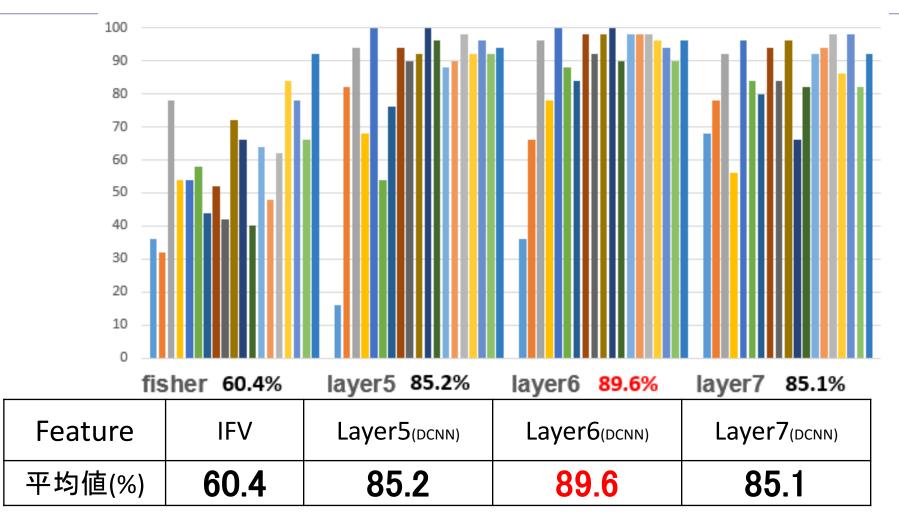
精度(平均適合率):89.6%

・ オノマトペ画像の認識(20種類 各50枚)





# 自動収集されたオノマトペ画像 の精度評価



リランキングにはLayer6が有効



#### 報告内容(2)

- 質感画像データセット自動構築システム
- ・Web収集画像を用いた質感語の視覚性評価 実験.
  - オノマトペに対応する画像の認識可能性評価
  - 122オノマトペ認識可能性ランキング.
  - 同一名詞内での、オノマトペの分類可能性の評価
- ・ CNN特徴 (feature map) を用いた質感領域抽出

## Web収集画像を用いた<sup>電気</sup> オノマトペの「視覚性」評価

「ごちゃごちゃ」「ぐるぐる」「ごつごつ」 「ざらざら」「ぐつぐつ」....

どれが画像に表れやすくて、 どれが画像に表れにくい?

視覚的なオノマトペはどれ?

オノマトペの深い理解を目指す!

オノマトペと画像の関係の理解.

画像からのオノマトペの自動生成!?









© 2016 UEC Tokyo.



#### 認識可能性の評価

- ・測定方法
  - ランダム画像5000枚を 正例画像50枚に混ぜ、 学習モデルを使って、分離 (5 fold cross validation)
  - 一分離度合いで、認識可能性を評価

分離度大⇒認識可能性が高い 分離度小⇒認識可能性が低い





#### 画像特徴量・分類器

- ・画像特徴量(2種類を比較)
  - Improved fisher vector (IFV)
    従来手法
    - SURF 128次元、クラスタ数 256
  - Deep Convolutional Neural Network (DCNN)

深層学習による特徴量

- ・分類器
  - Support vector machine (SVM)
    - ・線形SVM



#### DCNN特徴の抽出

- Overfeat [P. Sermanet et al., 2013]
  - ImageNet Challenge の 1000 カテゴリ, 100万枚でpre-training すみ.
  - layer5, 6, 7 の出力結果を扱う

Layer5: conv

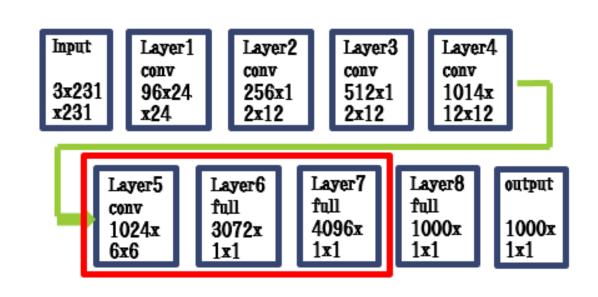
36864 次元

Layer6: full

3072 次元

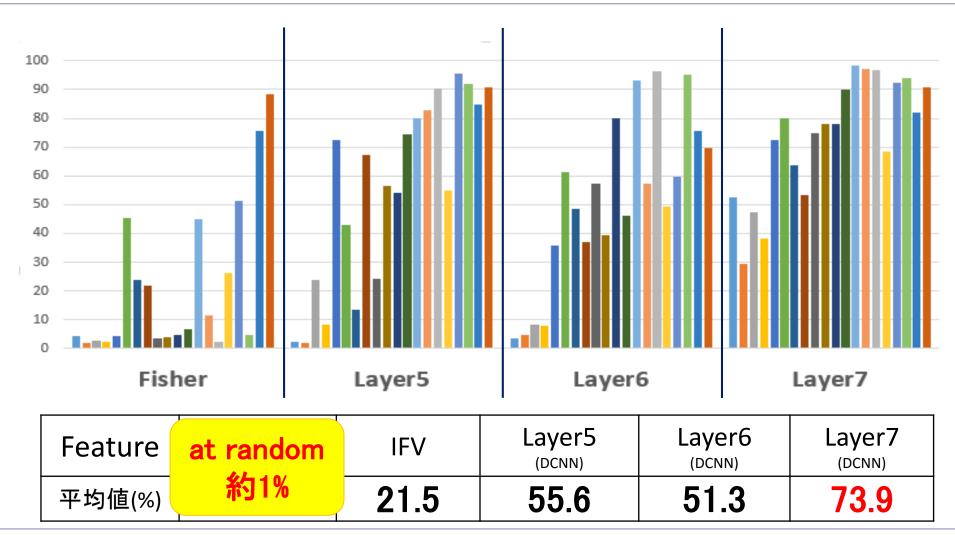
Layer7: full

4096 次元



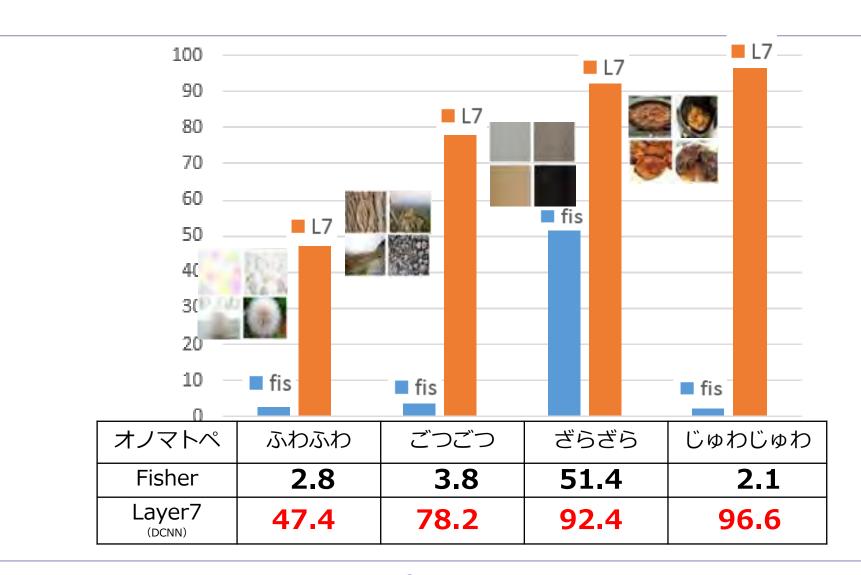


## 実験(2): 認識可能性の評価・ランダム画像5000枚との分離度





#### 認識可能性の評価例(4種類)





## ランダム画像との分離結果 (1)

#### 「ふわふわ」(赤枠はノイズ画像)

Fisher 2.8%

Layer7 (DCNN) 47.4%





#### 「ごつごつ」

Fisher 3.8%

Layer7(DCNN) **72.8%** 







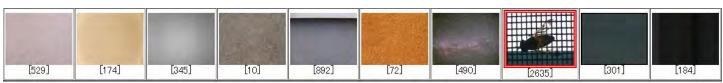
## ランダム画像との分離結果(2)

#### 「ざらざら」(赤枠はノイズ画像)

**Fisher 51.4%** 



Layer7 (DCNN) 92.4%



#### 「じゅわじゅわ」

**Fisher** 

2.1%



Layer7(DCNN)

96.6%



# オノマトペ122画像認識可能性 (視覚性) ランキング

#### B班 坂本先生 の オノマトペリストを利用して収集

- ・ 1 ふさふさ 100.0
- ・ 2 ふくふく 100.0 (
- ・ 3 べとべと 99.6
- ・ 4 ぶつぶつ 98.7
- ・ 5 ぶくぶく 98.6
- ・ 6 しゃきしゃき 98.4
- ・ 7 ざらざら 96.7
- ・8 つやつや 96.7
- ・ 9 もふもふ 96.4
- 10 くたくた 96.1()

http://mm.cs.uec.ac

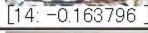


- ・ 11ぱさぱき
- ・ 12くるくる
- ・ 13さくさく
- ・ 14ぎしぎし
- 15ねばね [203: -0.269475]
- 16ゆらゆき
- 17さらさら
- ・ 18ぼさぼむ
- ・ 19かちかれ
- ・ 20ごてご



























[42: -0.308355

[4: -0.375956 ]



[35: -0.382593] [379: -0.384824]















## オノマトペ視覚性ランキング 下位20

- ・ -20 ふにゃ
- -19 もさも
- ・ -18 ぐにゅ
- -17 がしか
- -16 するす
- ・ -15 しとし
- ・ -14 もわも
- -13 じゃり
- ・ -12 とげと
- ・ -11 ごろこ























[74: -0.742678]



551: -0.738117









509: -0.7567

909: -0.757506

## オノマトペ視覚性ランキッグ

#### 下げつり

- -10 がさがさ
- -9 こなこな
- -8 ねちょね [150: -0.66234
- -7 うにょうり
- -6 かくかく
- -5 かしかし
- -4 てらてら
- -3 ばしばし
- -2 ずしずし
- -1 うぞうぞ









































#### 同一種物体内のオノマトペ分類





「じゅわじゅわ」画像 = 食事画像 (?)



「ごつごつ」画像 = 山の画像 (?)

特定の物体に偏る

質感認識ではなく、物体認識をしている可能性...

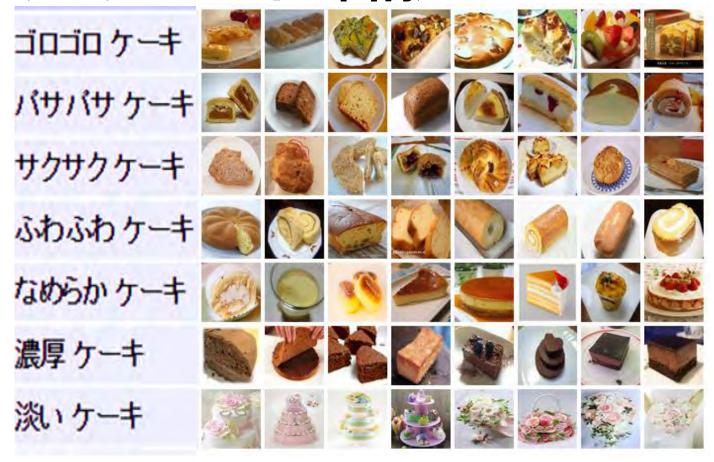


名詞+オノマトペで同一名詞に関連した画像 を収集 ⇒ オノマトペのマルチクラス分類

物体クラスの違いによる分類の可能性を排除する

### 名詞+オノマトペ画像の認識

ケーキ+オノマトペ画像



## 同一名詞内オノマトペ画像<sup>電気通信大学</sup>マルチクラス分類

- DCNN特徴 5 fold cross varidation で実験
- 名詞と組み合わせたオノマトペは、Webテキスト検索の共起頻度に基づき自動選択
  - クラス数が少なくなるので形容詞も一部使用

6~8クラス (300~400枚)

	犬	靴	ケーキ	(300°4
名詞	8クラス	6クラス	7クラス	7クラス
分類率(%)	52.5	85.7	72.3	84.6

#### ケーキ内オノマトペマルチクラス分類

				00	nfusio	n mat	rix		
ゴロゴロ ケーキ		31	3	10	2	3	1	0	62.0
バサバサ ケーキ		3	26	3	7	5	6	0	52.0
サクサクケーキ	664	8	4	24	8	2	3	1	48.0
ふわふわケーキ		1	2	2	42	3	0	0	84.0
なめらか ケーキ		3	2	3	3	37	2		61.5%
濃厚 ケーキ		1	0	2	1	0	46	0	92.0
淡いケーキ		0	0	0	0	3	0	47	94.0
		66.0	70.3	54.5	66.7	69.8	79.3	97.9	72.3%



#### 名詞+オノマトペ画像の認識

・ 花 + オ / マトペ 画像 ポンポン 花 ふわふわ 花 フレッシュ 花 メイン 花 ブルー 花 黄色い 花 赤い花

### 名詞内オノマトペマルチクラ人 分類

		confusion matrix							
ポンポン 花 🥻		35	3	2	2	4	0	4	70.0
ふわふわ 花 🧖	and to the	7	36	2	3	2	0	0	72.0
フレッシュ 花 🥻	A PAPER	2	1	38	7	1	0	1	76.0
メイン花	多多	0	2	3	44	1	0	0	88.0
ブルー花 🥻	多級	1	0	0	0	49	0	0	98.0
黄色い 花	* 3	0	0	0	0	0	50	0	100.0
赤い花	<b>1</b> 1	3	0	0	0	3	0	44	88.0
		72.9	85.7	84.4	78.6	81.7	100.0	89.8	84.6%

## 弱教師あり学習による 質感画像領域分割分析

#### 領域分割学習データ



### 構築の困難性

- ・ 大量の学習サンプルが必要
  - 数万枚から数百万枚
- 領域分割(応用)においてはピクセルレベルのアノテーションが必要
  - コストが爆発的に増加

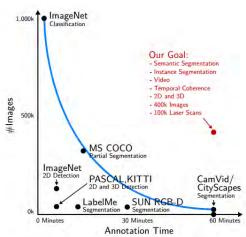


画像認識の アノテーション



領域分割のアノテーション





[J. xie+, CVPR2016]

## 弱教師あり学習による<sup>戦闘</sup> 領域分割

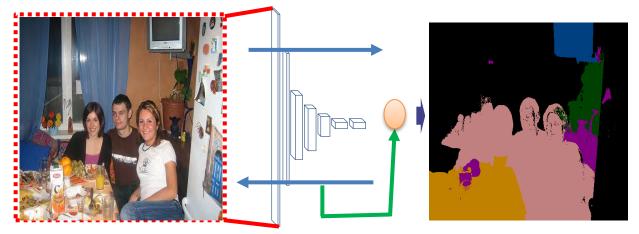
- ・弱教師あり学習による領域分割
  - ラベル情報のみで学習を行い、領域分割
- アノテーションのコストを大きく抑える





# 手法③ 伝搬値の差分による 高速かつ高精度な領域分割

- ・ 伝搬値の勾配のみを使って領域分割
- ・高速かつ高精度に



詳細な位置の推定

伝搬値から各カテゴリの

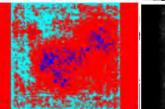
W. Shimoda and K. Yanai: Distinct Class-specific Saliency Maps for Weakly Supervised Semantic Segmentation, **ECCV 2016**,

#### BPによる認識結果の可視化

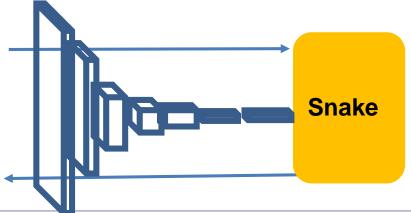
- 学習プロセスにおけるBack propagation(BP)を 利用して、認識結果の可視化が可能
  - 学習に用いる勾配の大きさが物体の位置に反応する
  - 認識結果の可視化→物体の位置の推定に
  - 弱教師あり学習手法
  - この手法をベースに

[K. Simonyan+, ICLR 2014]





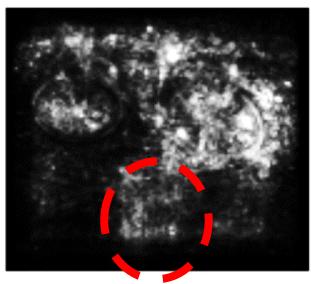




### Simonyanの手法の問題点

- ・関係のない場所が反応
- ・ 複数の物体が画像にうつっている場合に精度 が極端に下がる





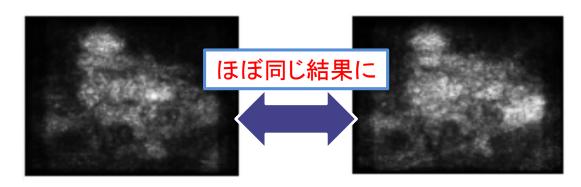


#### 問題点

- Back propagationから得られる勾配のみで完 結をさせたい
- Simonyanの手法はSemanticな情報が失われて しまう



対象画像

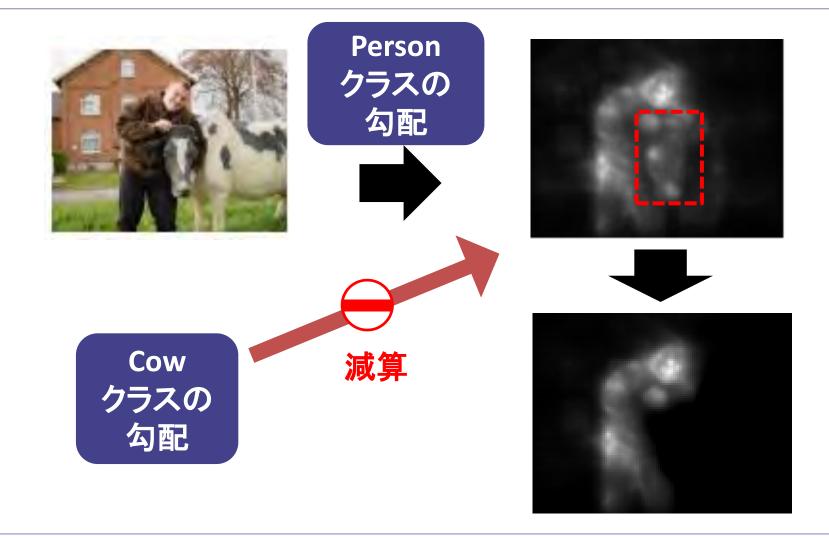


Personクラスから 得られる勾配

Trainクラスから 得られる勾配



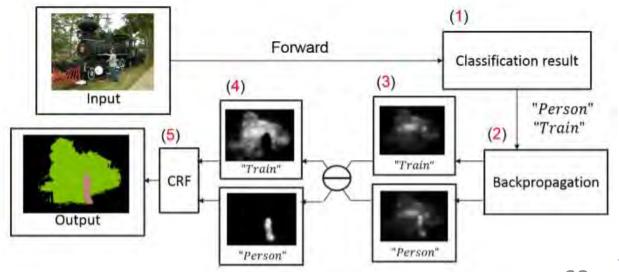
#### 勾配についての差分





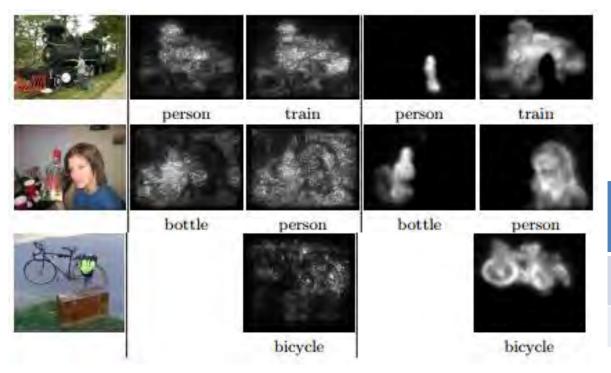
#### 手法の詳細

- 1. 画像にうつっている物体の認識
- 2. 認識した領域について、Back propagation
- 3. 各クラスの勾配について互いに減算
- 4. CRFを使って領域を統合





### Simonyan et al. との比較



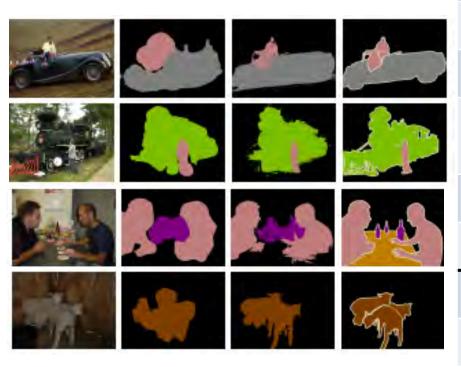
Method	Mean IOU
Sim et al. + CRF	33.8
Ours	44.1

Simonyan et al.

**Ours** 



#### 結果の例と最新手法との比較



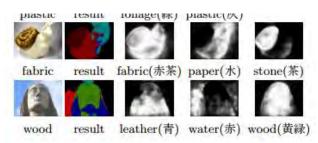
Method	Mean IOU
MIL-FCN (iclr 2015)	25.7
EM-Adapt(iccv 2015)	38.2
<b>CCNN</b> (iccv 2015)	34.5
MIL-sppxl (cvpr2015)*	36.6
MIL-bb (cvpr2015)*	37.8
MIL-seg (cvpr2015)*	42.0
Ours w/o CRF	40.5
Ours w/ CRF	<u>44.1</u>

\* は追加画像を用いている手法

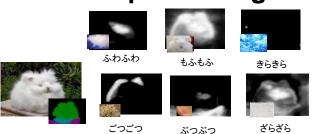
#### 弱教師あり学習領域分割の応用

質感画像など、アノテーションの困難な対象のデータについても有効

#### **Material images**



#### **Onomatopoeia images**



#### **Food images**



#### Satelite images (in AIST)







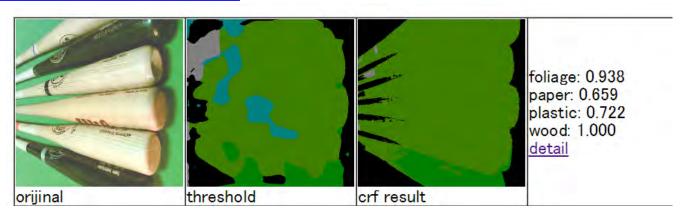


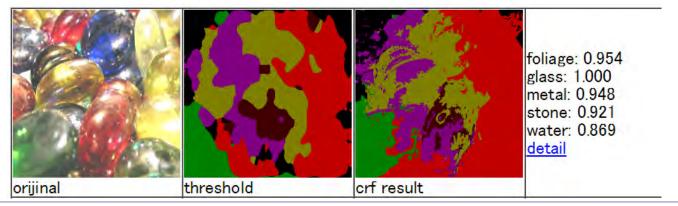




#### FMDの領域分割

http://mm.cs.uec.ac.jp/shimoda k/space/mat/fmd/index.cgi?EID=2&menu=0&1PG=4&2PG=0&3PG=0&CATE=
 0&NID=1&GPID=6&TID=1

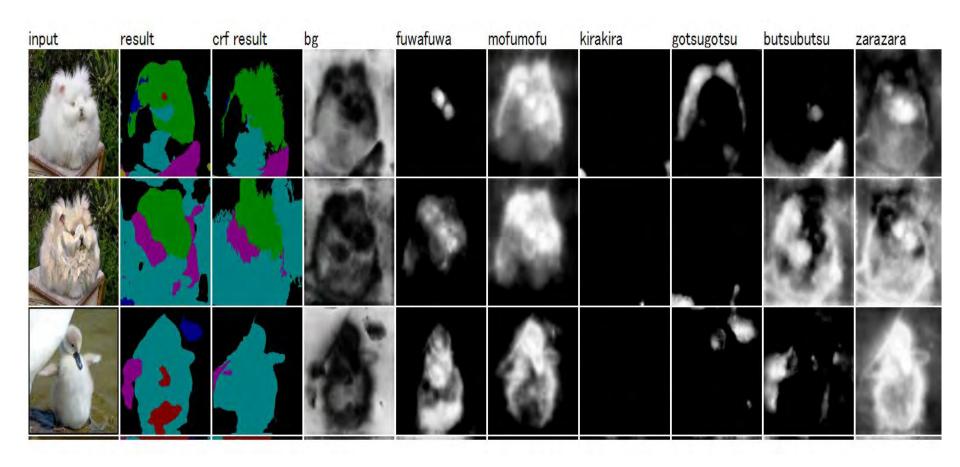






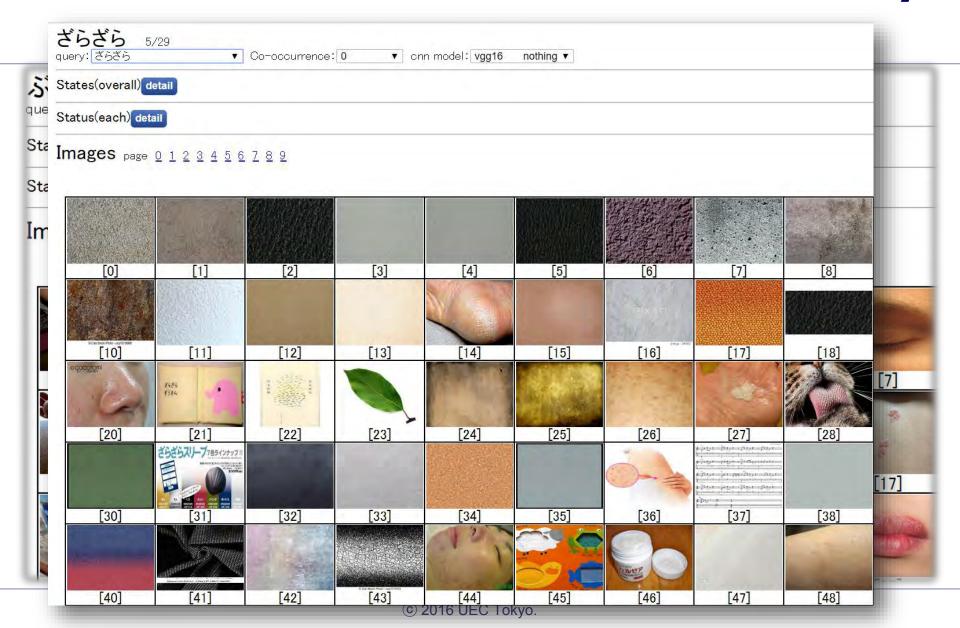
## オノマトペの領域分割

http://mm.cs.uec.ac.jp/shimoda-k/space/caffe9/caffe-master/models/ono/





#### Web画像による学習データが too noisy



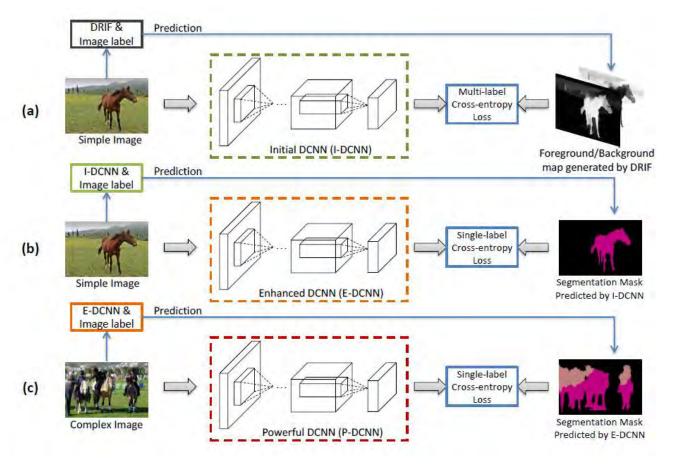


## さらなる性能向上弱教師有手法

- ・完全教師有手法の利用.
  - 弱教師有手法の結果を使って、完全教師有手法を学習
  - 学習⇒認識⇒学習データの更新 を繰り返す.

- ・さらに、領域分割の容易さを評価
  - 簡単なデータから学習する
- ・容易な画像でデータ拡張

# Simple-To-Complex framework



Y. Wei, X. Liang, Y. Chen, et al.: STC: A Simple to Complex Framework for Weakly-supervised Semantic Segmentation, arXiv:1509.03150, (2015).



### 最新結果: state-of-the-art

Methods	bg	aero	bike	bird	boat	bottle	pns	car	cat	chair	woo	table	gop	horse	motor	person	plant	sheep	sofa	train	14	Molm
Fully Supervised:																					0.71	
O2P [5]	85.4	169.7	22.3	45.2	44.4	49.6	66.7	757.8	56.2	13.5	46.1	32.3	341.2	259.1	55.3	51.0	36.2	50.4	27.8	46.5	44.6	47.6
SDS [10]	86.	3 63.3	25.7	63.0	39.8	59.2	70.9	61.4	54.9	16.8	45.0	48.2	2 50.5	51.0	57.7	63.3	31.8	58.7	31.2	255.	748.5	51.6
FCN-8s [15]	-	76.8	34.2	68.9	49.4	60.3	75.3	374.7	77.6	21.4	62.5	46.8	871.8	63.9	76.5	73.9	45.2	72.4	37.4	170.9	55.1	62.2
Using Additional Supervision:																					10.1	
CCNN w/size [18]		42.3	24.5	56.0	30.6	39.0	58.8	352.7	54.8	14.6	48.4	34.2	2 52.7	46.9	61.1	44.8	37.4	48.8	30.6	547.	741.7	45.1
One point[2]	80.0	5 50.2	23.9	38.4	33.1	38.5	52.0	50.9	55.4	18.3	38.2	37.7	751.0	146.1	54.7	43.2	35.4	45.1	33.0	149.6	540.0	43.6
F/B prior + CheckMask[21]	87.	165.7	26.0	64.2	43.7	53.2	72.6	63.6	59.5	17.1	48.0	43.7	761.2	252.0	69.3	54.8	43.0	50.3	34.6	559.	242.0	52.5
Weakly Supervised:																						
MIL-FCN [19]						-	-	-		-		-	-			-						24.5
EM-Adapt [17]	76.	3 37.1	21.9	41.6	26.1	38.5	50.8	344.9	48.9	16.7	40.8	29.4	447.1	45.8	54.8	28.2	30.0	44.0	29.2	234.3	3 46.0	39.6
CCNN [18]	1.	21.3	17.7	22.8	17.9	38.3	51.3	343.9	51.4	15.6	38.4	17.4	146.5	38.6	53.3	40.6	34.3	36.8	20.1	32.5	38.0	35.5
MIL-ILP-seg [20]	78.	7 48.0	21.2	31.1	28.4	35.1	51.4	155.5	52.8	7.8	56.2	19.9	53.8	3 50.3	3 40.0	38.6	27.8	51.8	24.7	133.	3 46.3	40.6
DCSM w/ CRF [28]	78.	1 43.8	26.3	49.8	19.5	40.3	61.6	53.9	52.7	13.7	47.3	34.8	3 50.3	48.9	69.0	49.7	38.4	57.1	34.0	38.0	40.0	45.1
F/B prior[21]	80.3	3 57.5	24.1	66.9	31.7	43.0	67.5	48.6	56.7	12.6	50.9	42.6	5 59.4	52.9	65.0	44.8	41.3	51.1	33.7	144.4	133.2	48.0
STC [29]	85.	2 62.7	21.1	58.0	31.4	55.0	68.8	63.9	63.7	14.2	57.6	28.3	3 63.0	159.8	67.6	61.7	42.9	61.0	23.7	252.	133.1	51.2
SEC [13]	83.5	5 56.4	28.5	64.1	23.6	46.5	70.6	558.5	71.3	23.2	54.0	28.0	068.1	62.1	70.0	55.0	38.4	58.0	39.9	38.	48.3	51.7
Ours	82.	7 63.5	28.5	60.2	29.7	60.0	71.6	61.9	62.9	19.6	45.8	47.5	5 47.0	56.1	70.6	58.3	31.7	62.4	38.7	36.5	38.3	51.2
Ours + Data Augmentation	83.0	67.5	29.7	69.7	28.8	59.7	71.2	266.4	69.8	18.6	49.8	44.7	7 49 4	60.5	73.5	61.8	32.7	62.7	39.0	34.3	3 36.5	52.8

## 画像の質感変化

## 深層学習は認識だけではない。 画像生成,画像変換も可能

①画像から 記号(言語)

> 画像認識 encoder

②記号(言語)「池のほとりでから画像 寝そべった」

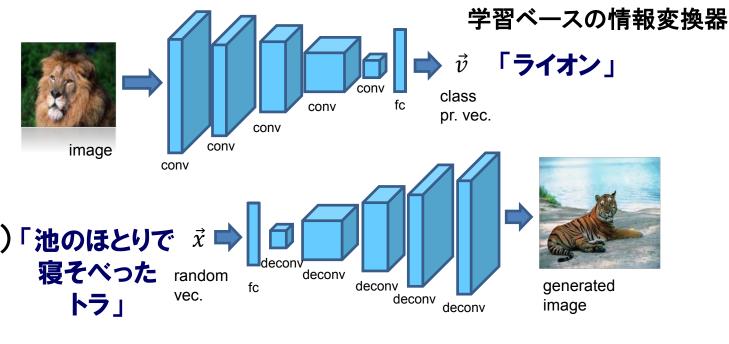
画像生成

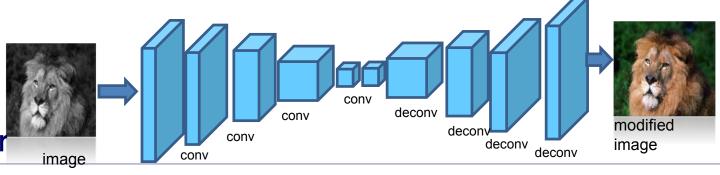
decoder

③画像から画像

画像変換

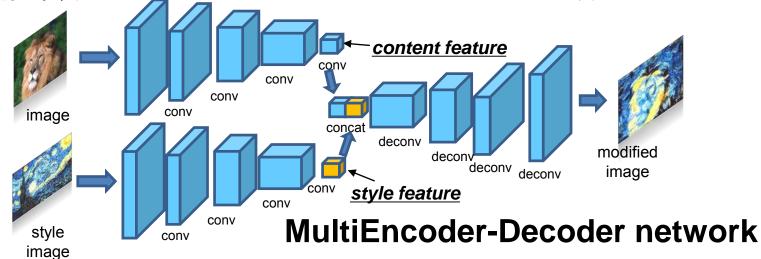
encoder-decoder



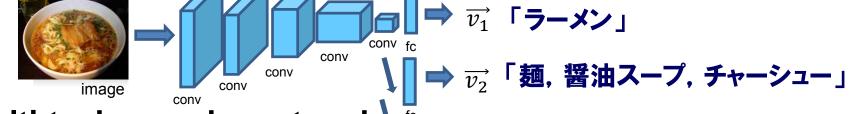


#### 複雑化するネットワーク

- 【研究者の仕事】
- ・ネットワーク設計
- ・誤差関数設計
- ・データセット構築
- ・ 複数の画像から1枚の画像を生成 (複数入力)



・ 1つの画像から複数の属性を同時推定(複数出力)



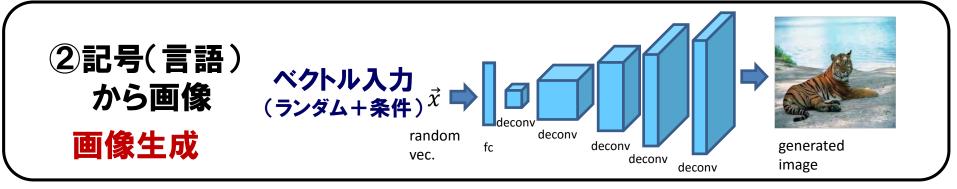
© 2016 EC Tokyo.

Multi-task encoder network  $\bigvee_{i=1}^{fc} \overrightarrow{v_3}$  650 kcal

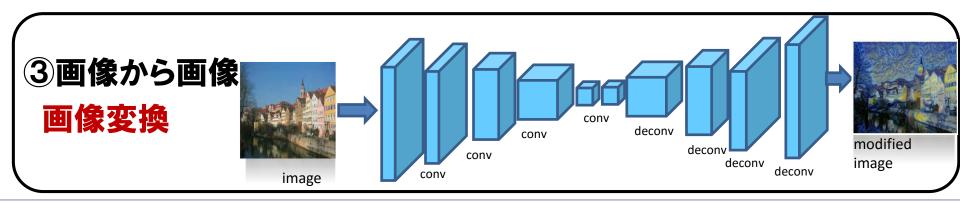


#### 深層学習で画像生成:2大手法

Generative Adversarial Network (GAN)



Neural Style Transfer





# Generative Adversarial Networks (GAN) (対立(敵対)ネットワーク)

#### 敵対する2つのネットワークを学習



学習済generatorに Zを連続変化させて与える.



近いZに対応する画像は似た画像になる. Zと画像の間に意味的な対応が自動的に学習できる.





Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. WardeFarley, S. Ozair, A. Courville, and Y. Bengio. Generative Adversarial Nets. NIPS 2014.

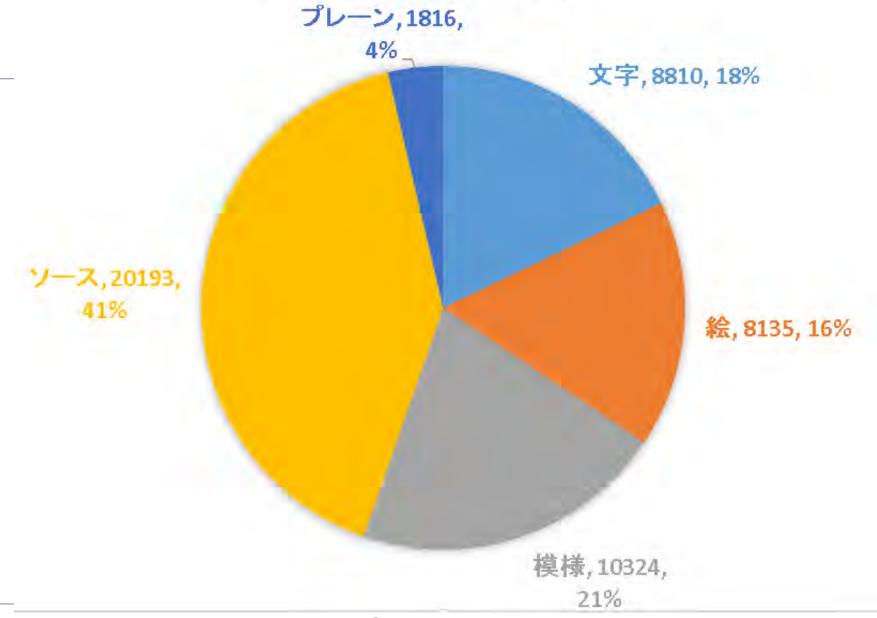
Emily Denton, Soumith Chintala, Arthur Szlam, Rob Fergus: Deep Generative Image Models using a Laplacian Pyramid of Adversarial Networks, NIPS 2015.

### 【余談】7万枚のラベル付きオムライズ画像 クラウドソーシング体験課題



83

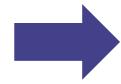
#### オムライス種別分布





### **Neural Style Transfer**





CNNを使って、画像を 様々な「スタイル」に変化 させる手法.



まさに「質感操作」!





#### 画像のスタイルの変換

- 「A Neural Algorithm of Artistic Style」
  - Gatys et al. Year 2015, arXiv:1508.06576
  - Deep Neural Networkを用いて、画像を絵画風に変換する
  - 元画像の構成を損なわずにスタイルを変換



コンテンツ画像

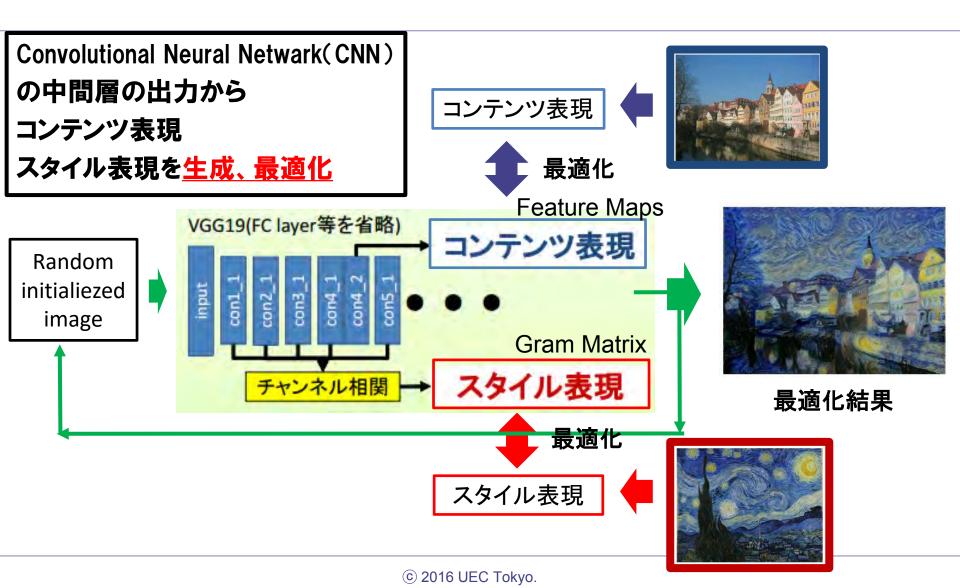


スタイル画像

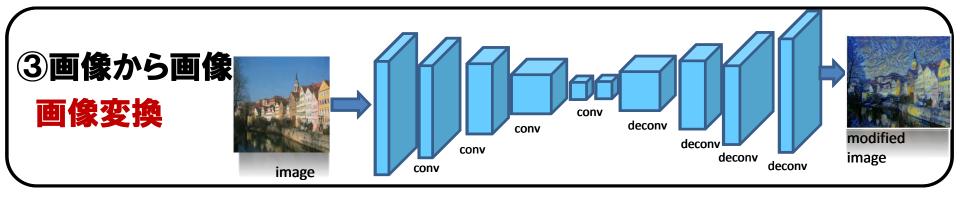




#### スタイルの変換の仕組み



# Encoder-Decoder net で よる高速変換



- ・画像変換:白黒⇒カラー化, 2D⇒3次元
- ・画像のスタイル変換
  - 質感変換
  - 「絵」のように変換

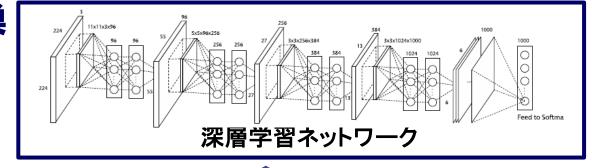
### 画像変換ネットの特徴:

### 逆変換ネット が 容易に構築可能

- ・入力、出力のペアが大量にあれば、
  - 一般には構築が難しい逆変換器を学習ベースで構築可能.正変換で学習データを大量自動生成.

(正変換は容易だが,逆変換は従来は不可能とされていた)

- 白黒⇒カラー変換
- 2D⇒3D変換
- 低解像度⇒高解像度変換(超解像度)



ビッグデータ + 巨大計算パワー

物理モデルは単なるデータ生成のための道具に...



#### 白黒 ⇒ カラー 変換

#### ・白黒⇒カラー変換

















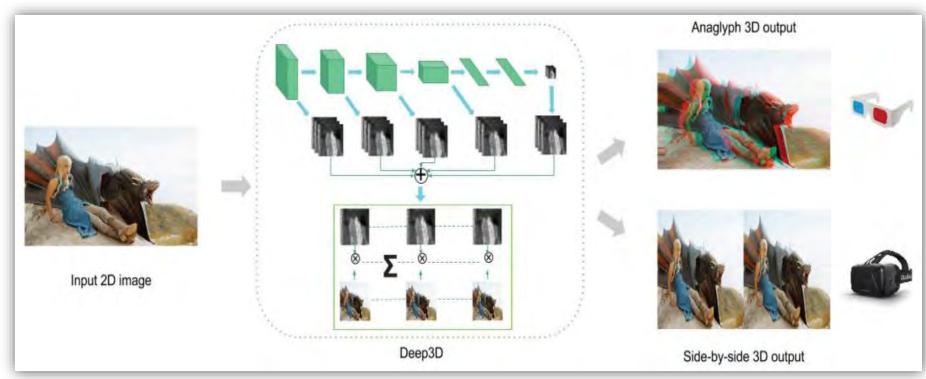


- •G. Larsson, M. Maire, and G. Shakhnarovich. <u>Learning Representations for Automatic Colorization</u>. ECCV 2016.
- •R. Zhang, P. Isola, and A. Efros. Colorful Image Colorization. In ECCV 2016.
- •S. lizuka, E. Simo-Serra, and H. Ishikawa. "Let there be Color!: Joint End-to-end Learning of Global and Local Image Priors for Automatic Image Colorization with Simultaneous Classification". SIGGRAPH 2016.



#### 2D ⇒ 3D変換

Deep 3D: 2D画像の自動 3D化

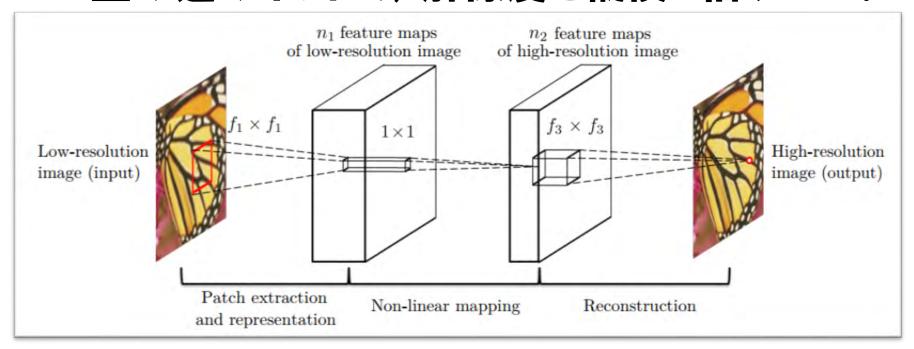


J. Xie, R. Girshick, and A. Farhadi: **Deep3D: Fully Automatic 2D-to-3D Video Conversion with Deep Convolutional Neural Networks, ECCV2016.** 



#### 超解像

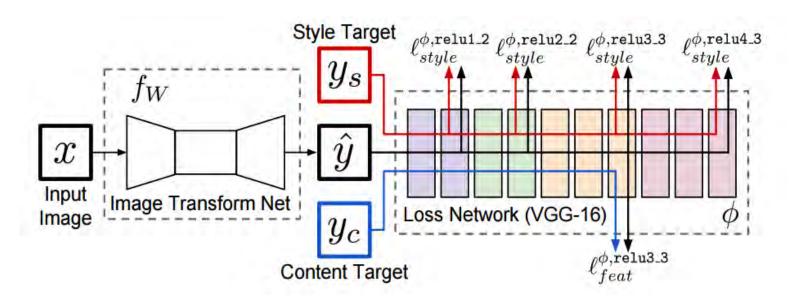
#### ・ 畳み込みネットで、解像度を縦横2倍ずつに、



Chao Dong, Chen Change Loy, Kaiming He, Xiaoou Tang. Learning a Deep Convolutional Network for Image Super-Resolution, in Proceedings of European Conference on Computer Vision (ECCV), 2014 W. Shi, J. Caballero, Ferenc Husz and J. Totz: "Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network", CVPR2016. 他多数

## Encoder-Decoder netで よる高速変換

Perceptual loss による Decoder-Encoder
 Networkの学習



J. Johnson, A. Alahi, Li Fei-Fei: Perceptual Losses for Real-Time Style Transfer and Super-Resolution, ECCV 2016



#### 食事画像(深大寺多聞そば)の質感変化



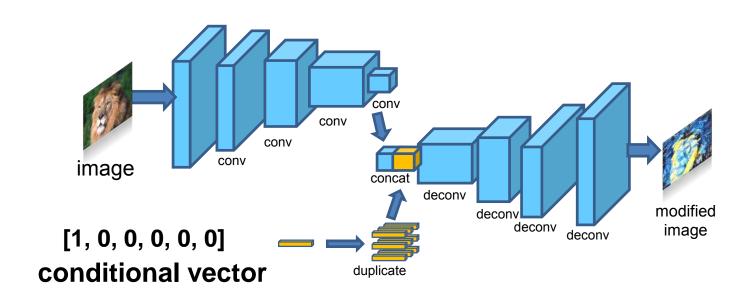






転送したスタイルの例

# Conditional vectorの追加 による複数スタイルの学習



Conditional Encoder-Decoder net. (Encoder-Decoder net. with CNN switch)



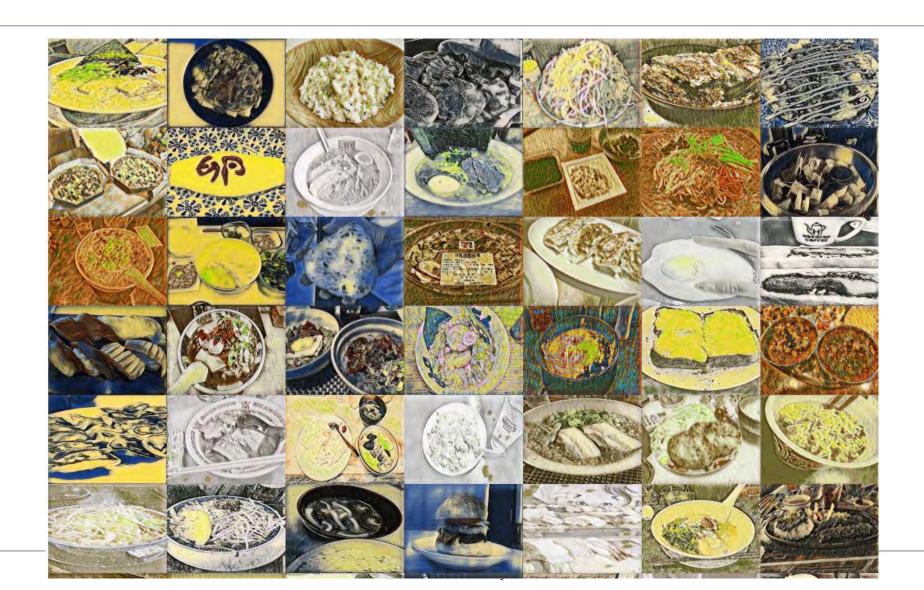
#### 高速マルチスタイル変換

#### 複数スタイルの融合 ⇒ デモあります!





## 食べ物は...





### Color-preserving !!!







食事画像認識/変換bot @fooding bot - 11月16日

[1]ラーメン(430kcal) [2]カツカレー(957kcal) [3]エビチリ(192kcal) [4]カ ツ丼(870kcal) [5]中華スープ(67kcal) @DADA\_Testcle





#### 食事画像認識/変換bot

@foodimg\_bot

Please send your meal photos to @foodimg\_bot. You can get category names and calorie intakes. Now the random neural style transfer engine also is working on !!

@ mm.cs.uec.ac.jp



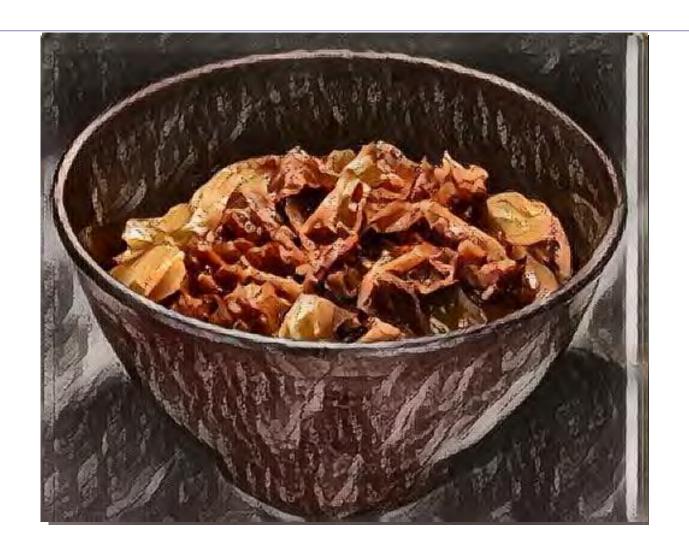
食事画像認識/変換bot @fooding bot - 11月16日

[1]天丼(560kcal) [2]たこ焼き(320kcal) [3]お好み焼き(550kcal) [4]納豆 (110kcal) [5]天ぷら盛り合わせ(410kcal) @maitake8918





## 牛丼!



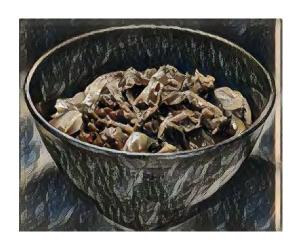


## ステーキも旨味アップ!





#### Which one you like to eat?















### Spatial style transfer

塗り分け可能. (unpublished)





### 全部ランダム!

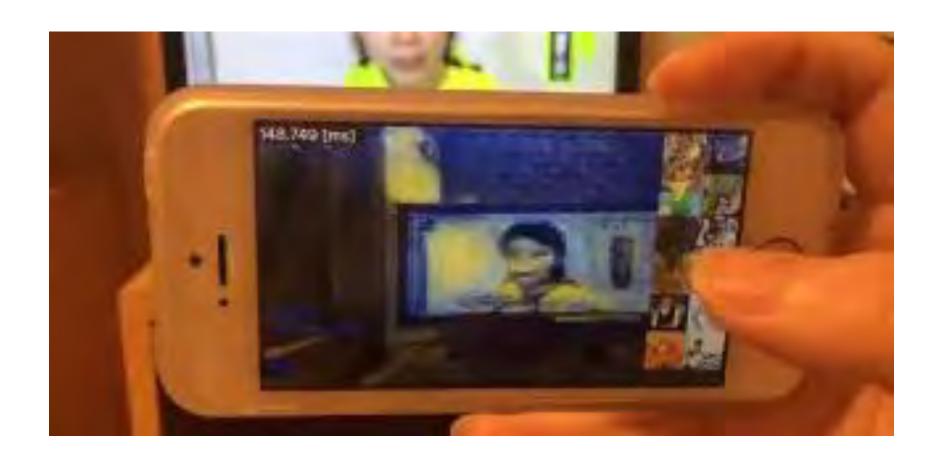
・16スタイル混合アート.







## モバイルdeep learning 応用: リアルタイムスタイル変換アプリ



#### この開発

#### RealTimeMultiStyleTransfer

デベロッパ: ryosuke tanno

App を購入、ダウンロードするには iTunes を開いてください。



#### iTunes で見る

この App は iPhone、iPad の両方に対応しています。

#### 無料

Available on the

**App Store** 

カテゴリ: 写真/ビデオ 更新: 2016年10月12日 バージョン: 1.1 サイズ: 14.2 MB 言語: 英語

販売元: ryosuke tanno © 2016 Yanai Lab.

#### 4+ 評価

互換性: iOS 9.2 以降。iPhone、iPad、および iPod touch に対応。

#### カスタマー評価

このアブリケーションの現行バー ジョンの平均評価を出すための十 分なデータがありません。

#### 説明

Recommend only for iPhone 7/6s/SE and iPad Pro.

This is an experimental application for academic research on mobile implementation of Deep Network.

RealTimeMultiStyleTransfer のサポート)

#### バージョン 1.1 の新機能

This version is a small improvement as follows.

· improve processing speed and image quality.

#### スクリーンショット



© 2016 UEC Tokyo.



#### 質感への応用

- Flickr material dataset(FMD)
  - 10種類の素材画像(布、金属、木製など)

- 素材画像の style transfer
  - 布を金属に
  - ガラスを布に





#### 質感画像におけるスタイル変換



#### スタイル変換+領域分割

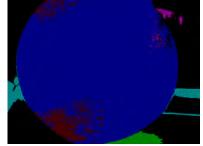
- ・問題点
  - 意図しない素材領域も変化・・・





- ・素材画像の領域を推定
  - 推定結果領域のみを変換可能!





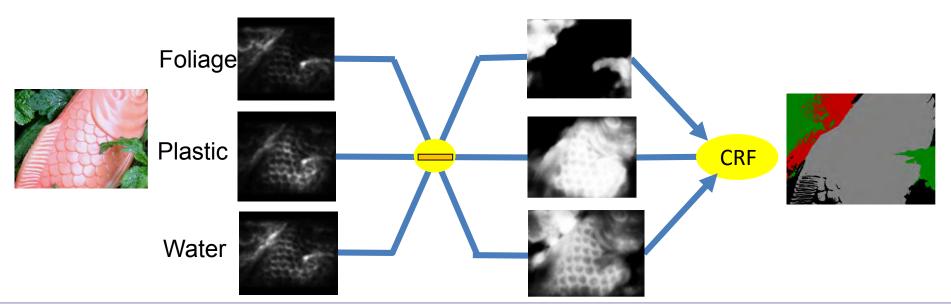


#### 質感画像の弱教師あり領域分割

Deep learningによる質感画像の認識



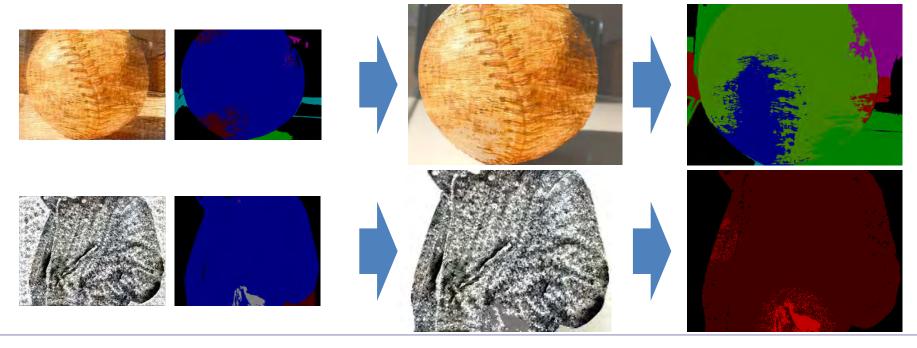
• 各カテゴリの認識に寄与した領域の推定



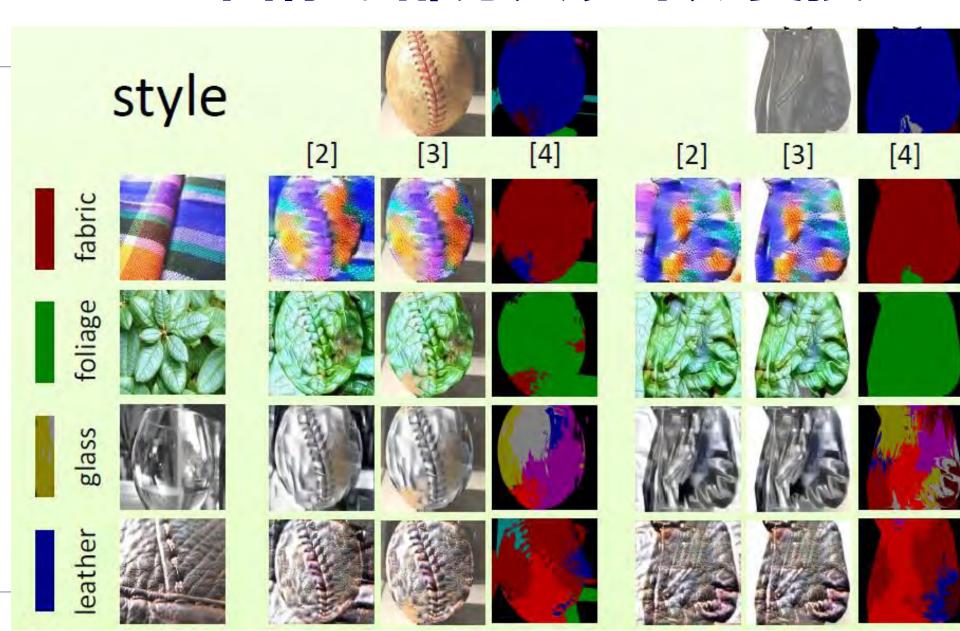


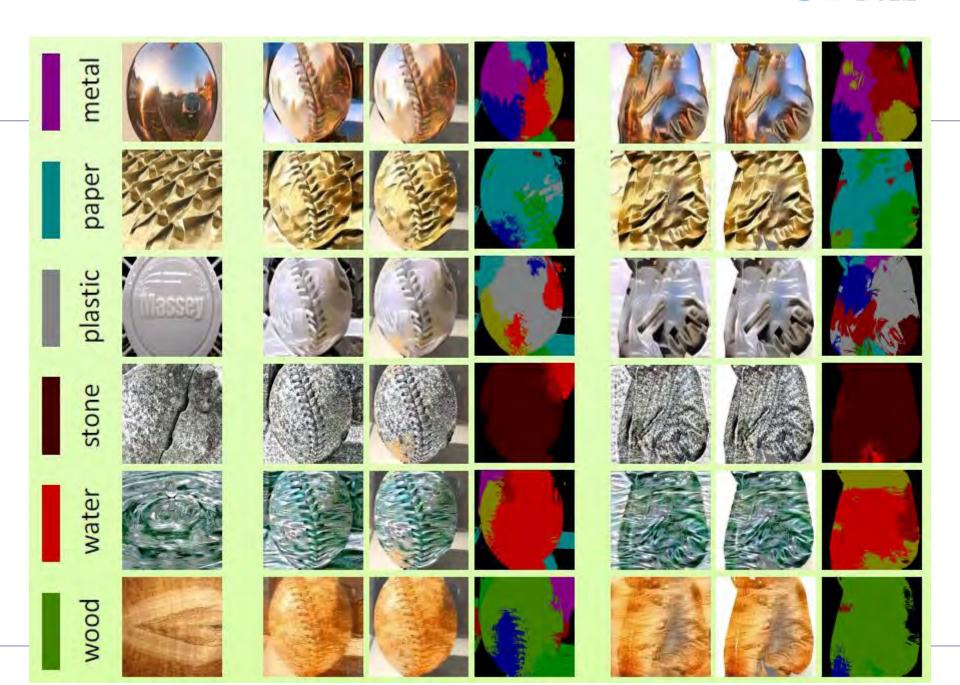
#### 部分的スタイル変換

- ・領域分割結果を用いて、対象物体のみを スタイル変換
- ・再度領域分割し、素材の変化の確認



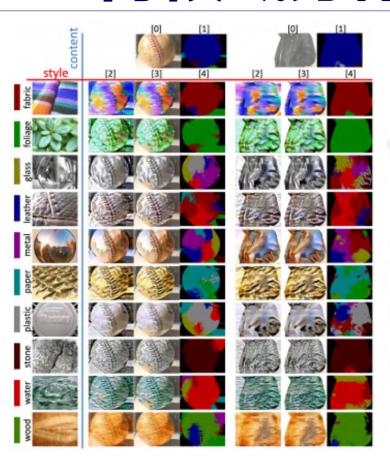
# FMD画像の部分スタイル変換







# 変換結果の一覧と再領域分割結果の評価



class	pixel acc.	mean IU	pixel acc.	mean IU				
fabric	0.83	0.72	0.96	0.93				
foliage	0.85	0.77	0.99	0.96				
glass	0.41	0.53	0.36	0.56				
leather	0.24	0.33	0.27	0.52				
metal	0.44	0.52	0.44	0.60				
paper	0.77	0.79	0.64	0.74				
plastic	0.60	0.60	0.54	0.67				
stone	0.87	0.79	0.95	0.93				
water	0.77	0.74	0.69	0.78				
wood	0.59	0.51	0.89	0.82				

- MetalやGlassのトランスファーが難しい
- ・ 反射の再現が困難?



#### オノマトペ画像の学習

- ふわふわ、もふもふ、きらきら、ごつごつ、ぶつ ぶつ、ざらざら
- 6種類、200枚ずつ

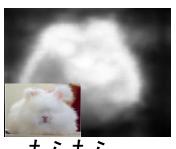




ふわふわ



ごつごつ



もふもふ



ぶつぶつ



きらきら

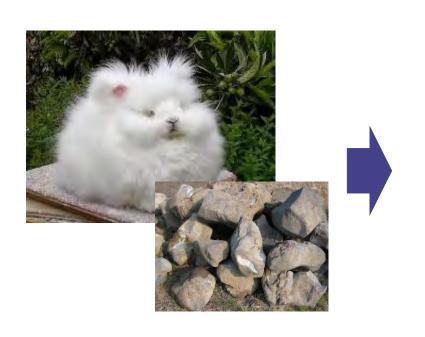


ざらざら



#### オノマトペのトランスファー

#### 「もふもふ」部分を「ごつごつ」にトランスファー!







## 言葉による 画像の任意質感変換

スタイル変換をより簡単に「ボールを金属化したい」と思ったも

Webでいい感じの金属スタイル画像を

探す必要がある









# Web画像マイニングを<sup>電気通信大学</sup>用いた任意質感変換



# 単語による画像収集と ウラスタリング

・ 質感単語で収集したWeb画像をスタイルの類似したもの同士のクラスタに分化



## 複数枚のスタイル画像が Gram Matrixの統合方法





style

content

mean maxiec Tokyo. minimize C



#### 質感単語によるスタイル変換

ごつごつ ふさふさ 質感単語 かちかち 自動構築した スタイルパッ スタイル変換 左:normal 右: color preserving



#### まとめ

- ・ Web画像データを利用した
  - オノマトペ画像認識と、認識可能性の評価
  - 言葉による任意質感変換の実現 について紹介した。

- 高速スタイルネットワークの改良による
  - -マルチスタイル,任意部分スタイル転送 について紹介した。

