

食事画像カロリー量推定における 回帰による手法と検索による手法の比較

會下拓実 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

食事画像カロリー量推定

Foodlog

料理をクロップ
料理の種類や量を**手入力選択**

1/4 人前	205 kcal
1/3 人前	279 kcal
1/2 人前	411 kcal
2/3 人前	551 kcal
3/4 人前	617 kcal
1 人前	823 kcal ✓
1.5 人前	1234 kcal
2 人前	1646 kcal

量の選択

カロナビ

食事写真を撮るだけで
カロリーがわかります

栄養士が実際にあなたの食事写真を見て
カロリーをチェックします

傾向が分かりやすい
体重グラフもあります

2015/3/1 Sun
1300 kcal

2014/10/8
54.0 kg

53.8 kg

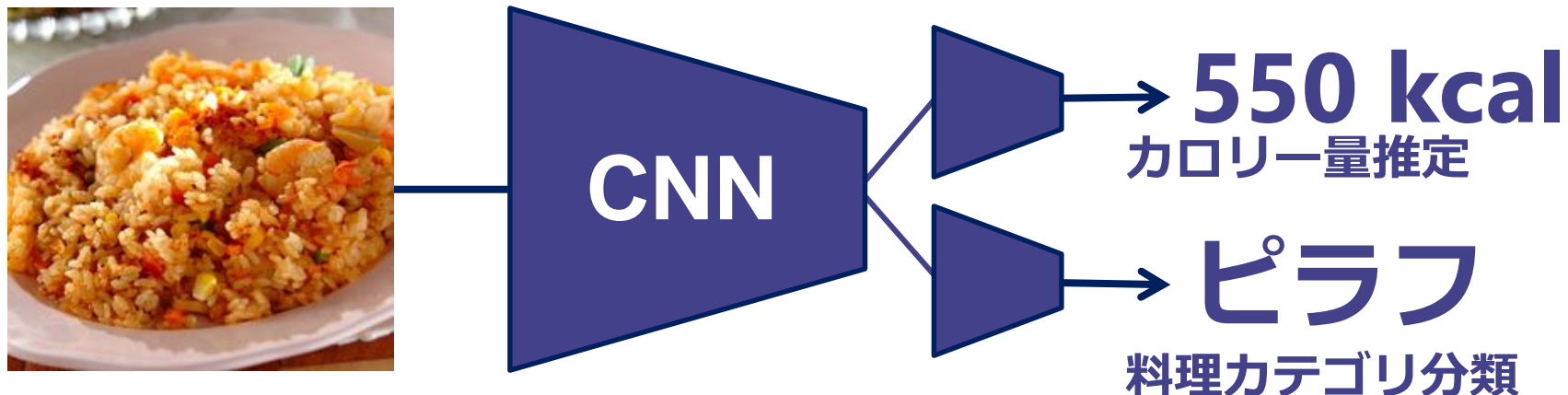
1ヶ月 3ヶ月 12ヶ月

栄養士による食事画像からの
カロリー量推定
有料サービス

食事画像からのカロリー量推定は未解決の問題

従来手法: 回帰による手法

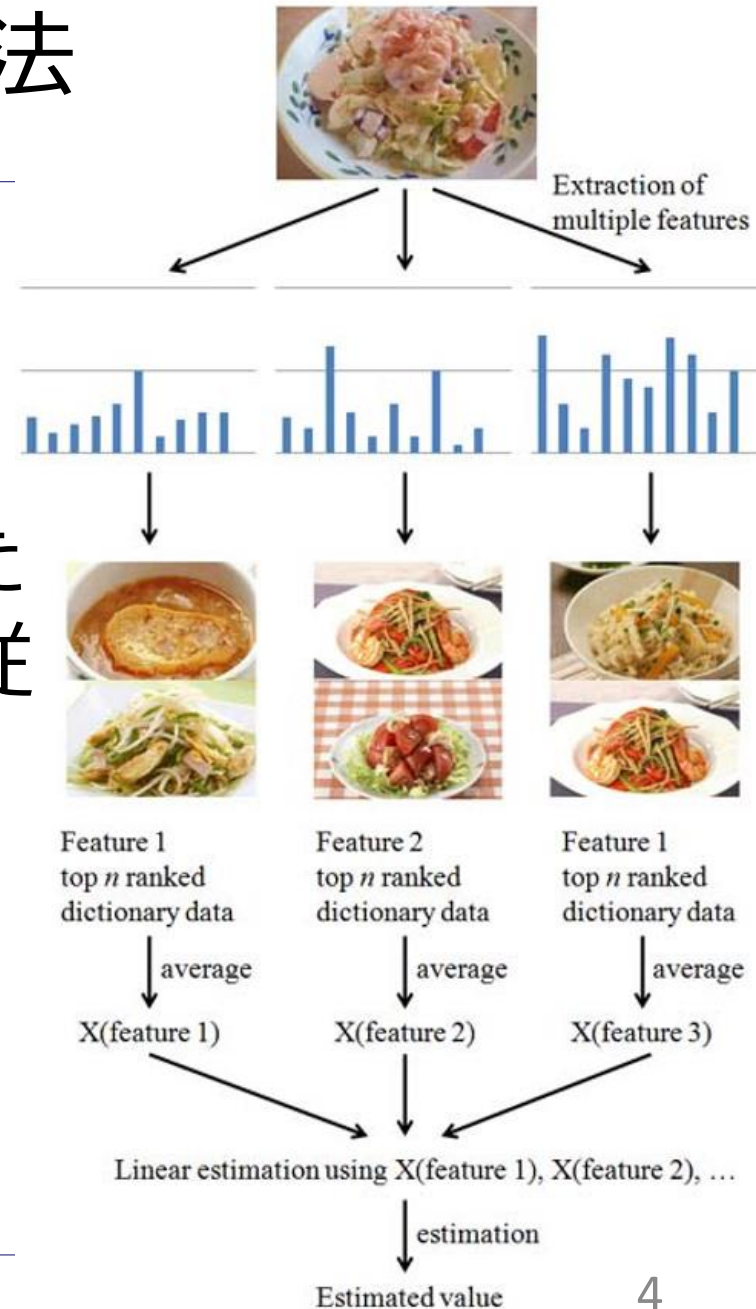
- 料理カテゴリ分類とカロリー量を同時に推定する Multi-task CNN.
- マルチタスク学習により精度向上



Ege and Yanai. Simultaneous estimation of food categories and calories with multi-task cnn. In Proc. of IAPR International Conference on Machine Vision Applications(MVA), 2017.

従来手法: 検索による手法

- 類似画像検索による
カロリー量推定
- 検索にはSURF特徴量を用いた
BoFやカラーヒストグラムなどの従
来特徴量を使用.

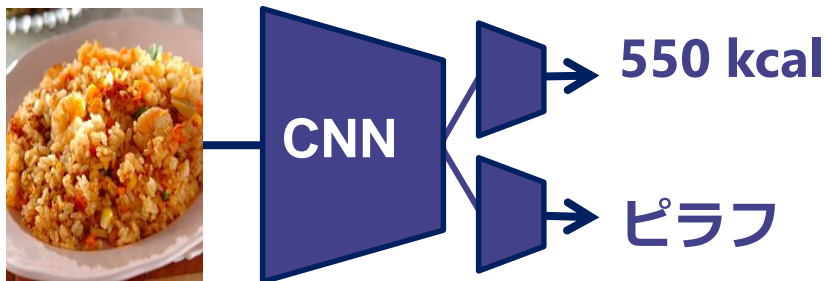


Miyazaki et al. Image-based Calorie Content Estimation for Dietary Assessment, Workshop on Multimedia for Cooking and Eating Activities, 2011.

二つの手法の比較

回帰による手法

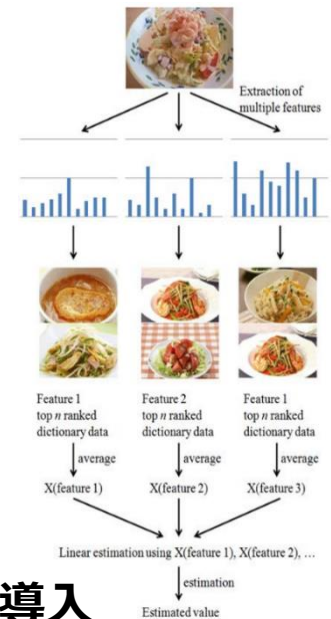
料理カテゴリ分類とカロリー量を同時に推定するMulti-task CNN. マルチタスク学習により精度向上.



検索による手法

類似画像検索による
カロリー量推定

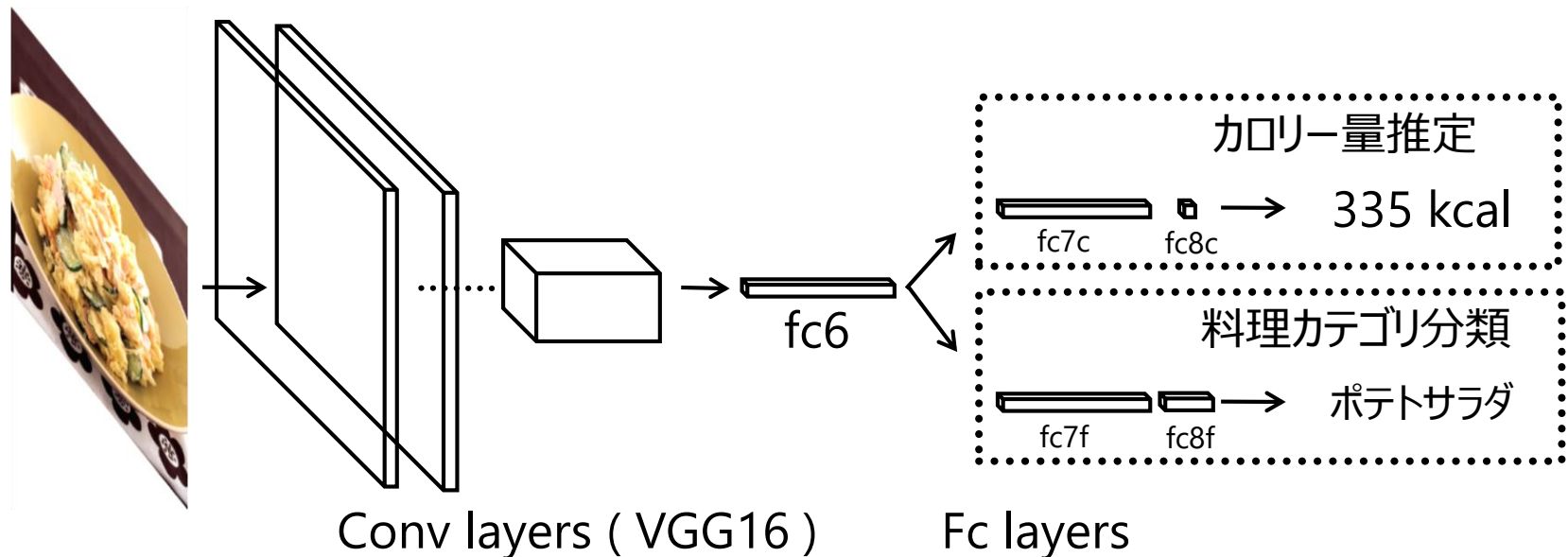
検索にはSURF特徴量を用いたBoFやカラーヒストグラムなどの従来特徴量を使用.



本研究では、最新の
CNNベースの距離学習を導入

回帰: Multi-task CNNの導入

- CNNを用いた食事画像からのカロリー量推定
- Multi-task CNNによるカロリー量と料理カテゴリの同時学習



Ege and Yanai. Simultaneous estimation of food categories and calories with multi-task cnn. In Proc. of IAPR International Conference on Machine Vision Applications(MVA), 2017.

回帰: Multi-task CNNの導入

- 損失関数はカロリー量推定ロス L_{cal} と料理カテゴリ分類ロス L_{cat} の線形和.
 - カロリー量推定ロス L_{cal} は相対誤差 L_{re} と絶対誤差 L_{ab} の線形和.

$$L_{ab} = |y_i - g_i| \quad L_{re} = \frac{|y_i - g_i|}{g_i}$$

y_i を画像 x_i から推定されたカロリー量とし,
 g_i を正解値とする.

- 料理カテゴリ分類ロス L_{cat} はソフトマックスクロスエントロピー.

検索: 画像検索によるカロリー量推定 CNNベースの距離学習の導入

画像検索によるカロリー量推定手順

- カロリー量付き食事画像データベースの構築.
- 画像特徴量に基づき類似画像を検索.
 - 本研究ではCNN特徴量を使用
- 検索された食事画像のカロリー量から推定.

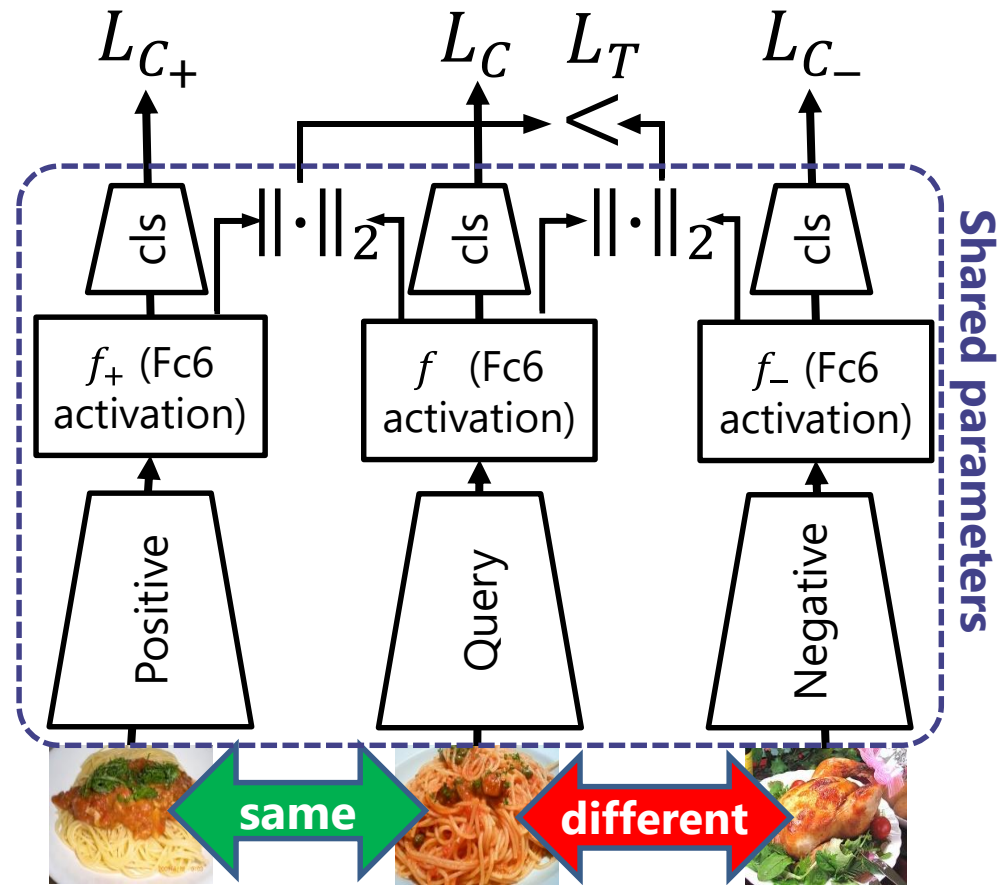


検索: 画像検索によるカロリー量推定

CNNベースの距離学習の導入

距離学習による画像特徴量の生成

- Triplet loss L_T とカテゴリ分類ロス L_{C_+} , L_C , L_{C_-} によりVGG16を学習.
- 全結合層の活性化値を画像特徴量として使用.



Triplet: J. Wangら, "Learning fine-grained image similarity with deep ranking," CVPR, 2014.

Triplet+class: E. Simo-Serraら, "Fashion style in 128 floats: Joint ranking and classification using weak data for feature extraction," CVPR, 2016.

検索: Triplet loss

- 損失関数は L_T と L_{C_+} , L_C , L_{C_-} の線形和.

- Triplet loss L_T

$$L_T = \max(0, g + \|f_+ - f\|_2 - \|f_- - f\|_2)$$

$\|f_+ - f\|_2$: 同じ料理カテゴリの画像のfc6活性値間のユークリッド距離.

$\|f_- - f\|_2$: 異なる料理カテゴリの画像のfc6活性値間のユークリッド距離..

- 料理カテゴリ分類ロス L_{C_+} , L_C , L_{C_-} はソフトマックスクロスエントロピー

カロリー量付き食事画像データセット

料理15カテゴリのカロリー量付き食事画像データセット. [Ege and Yanai, 2017]

レシピ情報サイト



みんなのきょうの料理 NHKエデュケーショナル



今日のレシピが必ず決まる!



約 **83,000** レシピ収集.



料理画像

調理時間 30分 🔥 エネルギー 310kcal
 塩分1.1g 野菜摂取量7g

※エネルギー・塩分・野菜摂取量

調理時間5分

じゃがいも・小	3個
ブロッコリー	1/4個
レモン・輪切り	4枚
パセリ・みじん切り	適量
小麦粉	適量
A「AJINOMOTO 胚芽の恵みコーン	大さじ1

カロリー量

- (1) さけは「コンソメ」をふって両面にひびませ、小麦粉をまぶす。
- (2) フライパンにAを熱し、(1)のさけの両面を中火で色よく焼き、弱火にしてフタをし、約3分蒸し焼きにする。
- (3) ジャがいもは皮をむいて3等分にし、水に10分ほどさらして水気をきる。鍋に入れ、ヒタヒタの水を加えて火にかけて、煮立ったら弱火にし、フタをしてやわらかくなるまで約10分ゆで、ザルに上げる。
- (4) 空鍋を火にかけて、(3)のじゃがいもを戻し入れ、鍋を揺すりながら粉をふかせて塩をふる。

カロリー量付き食事画像データセット

UEC food-100[1]の料理
100をアノテーション.

ノイズ除去.

サンプル数100以下の料理
カテゴリの画像を除去.

料理 **15** カテゴリ.

合計 **4877** 枚.



ピラフ



カレーライス



炒飯



焼きそば



スパゲッティ



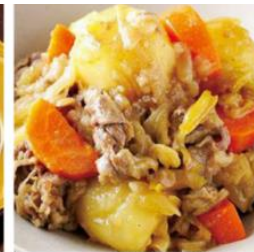
グラタン



味噌汁



シチュー



肉じゃが



ハンバーグ



冷奴



ちらし寿司



オムライス



ポテトサラダ



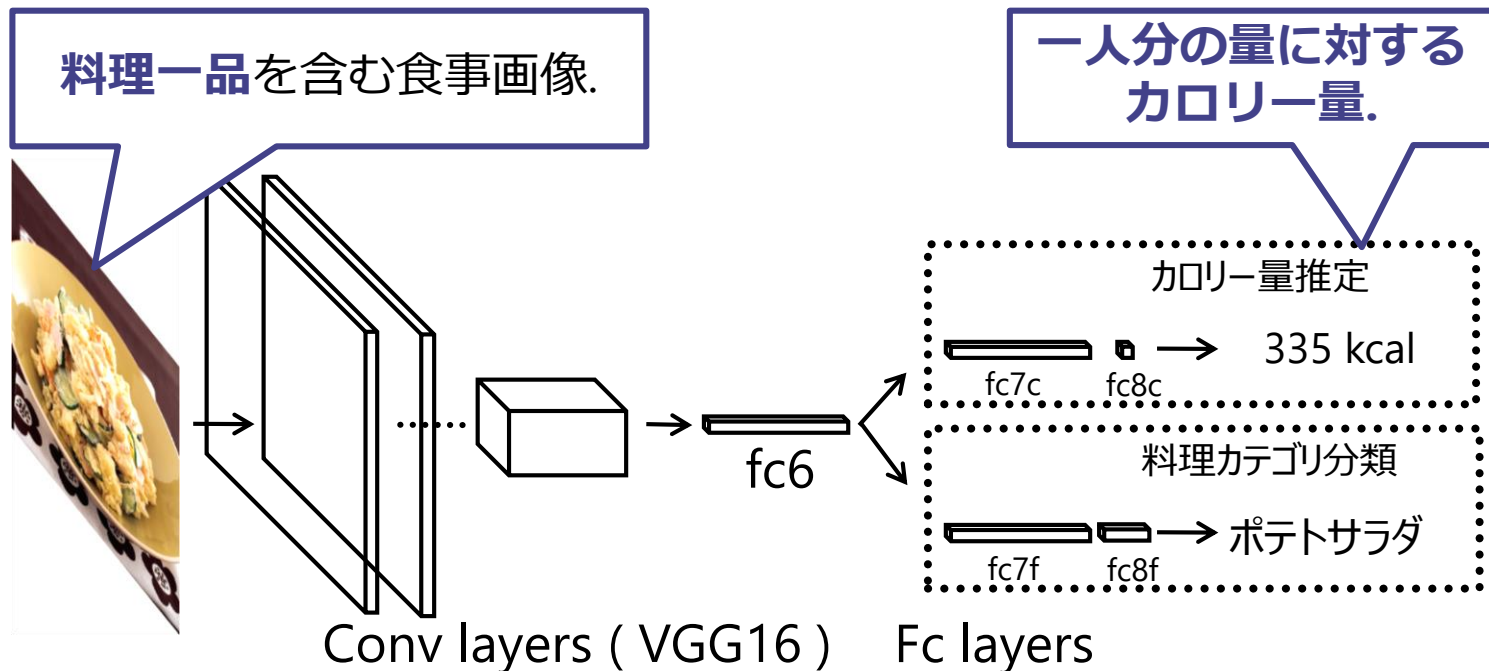
炊き込み
ご飯

[1] Matsuda et al. Recognition of multiple-food images by detecting candidate regions.
In Proc. of IEEE International Conference on Multimedia and Expo, 2012.

実験

実験：回帰によるカロリー量推定

- カロリー量付き食事画像データセットにより Multi-task CNNを学習.
 - 学習: 70%, テスト: 30%



実験：回帰によるカロリーー量推定

- カロリーー量推定評価指標
 - 相対誤差(%) : $|y_i - g_i|$
 - 絶対誤差(kcal) : $\frac{|y_i - g_i|}{g_i}$
 - 相関係数 : 推定カロリーー値と正解値の相関係数
 - < 20% (相対誤差)(%) : 相対誤差20%より小さい推定値の割合

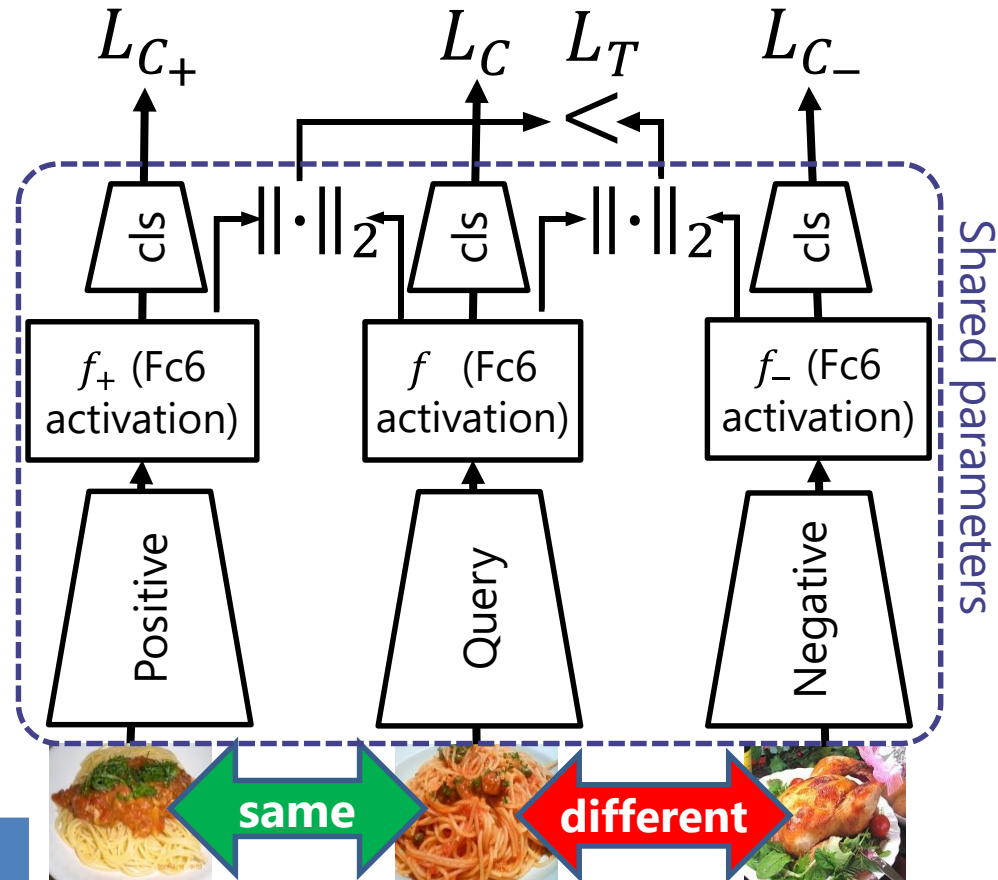
実験結果：回帰によるカロリー量推定

- VGG16, ResNet50/101に適用し,
マルチタスク学習による精度向上を確認。

		相対 誤差 (%)	絶対 誤差 (kcal)	相関 係数	< 20% (相対誤 差)(%)	クラス 分類 (%)
VGG16	VGG16 single-task	29.2	100.4	0.783	46.0	78.7
	VGG16 multi-task	28.0	96.5	0.805	47.2	83.2
ResNet50	ResNet50 single-task	31.1	96.7	0.789	47.0	83.7
	ResNet50 multi-task	30.6	93.2	0.803	49.0	83.9
ResNet101	ResNet101 single-task	29.4	94.7	0.797	49.4	83.8
	ResNet101 multi-task	29.6	91.9	0.807	50.8	82.0

実験：検索によるカロリー量推定

- Triplet loss と料理カテゴリ分類ロスをを用い, UEC food-100/256でVGG16を学習.
- 学習済みVGG16のfc6の活性化値を特徴ベクトルとして類似画像検索に使用.
- カロリー量付き食事画像データセットにより検索のためのデータベースを構築.
 - データベース : 70%,
テスト : 30%



料理一品を含む食事画像.

	相関係数	< 20% (相対誤差)(%)
従来特徴量の宮崎ら	0.32	35

予備実験：画像検索

- 各モデルから得られる特徴量による同カテゴリ画像検索
 - ① ImageNetにより事前学習済みのVGG16
 - ② UEC food-100/256により学習
 - ③ UEC food-100/256により **triplet loss** で学習
 - ④ UEC food-100/256により **triplet loss + クラス分類** で学習
- テストにはカロリーー量付き食事画像データセットを使用

		モデル	mAP@1	mAP@5	mAP@10
ImageNet pre-trainモデル	}	① fc6	39.8	48.7	46.2
		①+②(100) fc6	56.7	65.4	62.3
UEC food-100/256 fine-tuneモデル	}	①+②(100) fc7	57.8	65.7	62.4
		①+②(256) fc6	47.8	56.8	54.4
		①+②(256) fc7	47.3	56.3	54.2
Triplet loss fine-tuneモデル	}	①+③(100) fc6	41.2	50.8	48.5
		①+②(100)+③(100) fc6	57.7	65.5	62.9
Triplet loss + class fine-tuneモデル	}	①+④(100) fc6	65.8	72.1	69.8
		①+②(100)+④(100) fc6	61.3	68.6	66.1

実験結果：検索によるカロリー量推定

- 各モデルから得られる特徴量によるカロリー量推定
 - ① ImageNetにより事前学習済みのVGG16
 - ② UEC food-100/256により学習
 - ③ UEC food-100/256により **triplet loss** で学習
 - ④ UEC food-100/256により **triplet loss + クラス分類** で学習
- テストにはカロリー量付き食事画像データセットを使用

		モデル	相対誤差 (%)	絶対誤差 (kcal)	相関係数	< 20% (相対誤差)(%)
UEC food-100 fine-tuneモデル	{	①+②(100) fc6 k=5	37.6	103.2	0.748	46.4
		①+②(100) fc6 k=10	37.9	102.0	0.761	47.0
		①+②(100) fc6 k=15	38.0	101.0	0.764	47.2
Triplet loss fine-tuneモデル	{	①+③(100) fc6 k=15	48.8	114.5	0.685	44.4
		①+②(100)+③(100) fc6 k=15	38.4	101.5	0.754	45.9
Triplet loss + class fine-tuneモデル	{	①+④(100) fc6 k=15	34.0	94.5	0.783	49.5
		①+②(100)+④(100) fc6 k=15	36.7	98.6	0.777	48.1

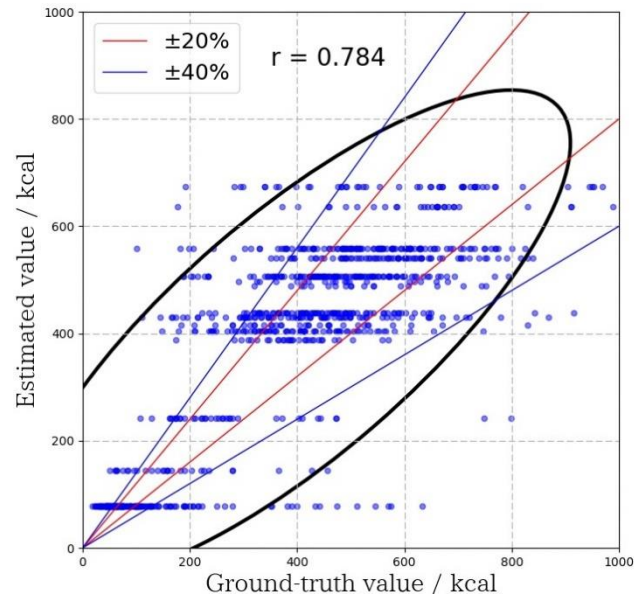
実験結果：手法の比較

- VGG16を用いた場合の比較.
- テストにはカロリー量付き食事画像データセットを使用.
- Baseline (クラス分類) : 料理カテゴリごとにカロリー値を固定.

		相対 誤差 (%)	絶対 誤差 (kcal)	相関 係数	< 20% (相対誤 差)(%)
固定	Baseline	32.4	93.6	0.784	50.0
回帰	VGG16 single-task	29.2	100.4	0.783	46.0
	VGG16 multi-task	28.0	96.5	0.805	47.2
*参考	*VGG16 multi-task(+ 食材, 調理手順)	27.4	91.2	0.817	50.1
検索	VGG16(UEC food-100)	38.0	101.0	0.764	47.2
	VGG16(ImageNet) triplet + class	34.0	94.5	0.783	49.5
	VGG16(UEC food-100) triplet + class	36.7	98.6	0.777	48.1

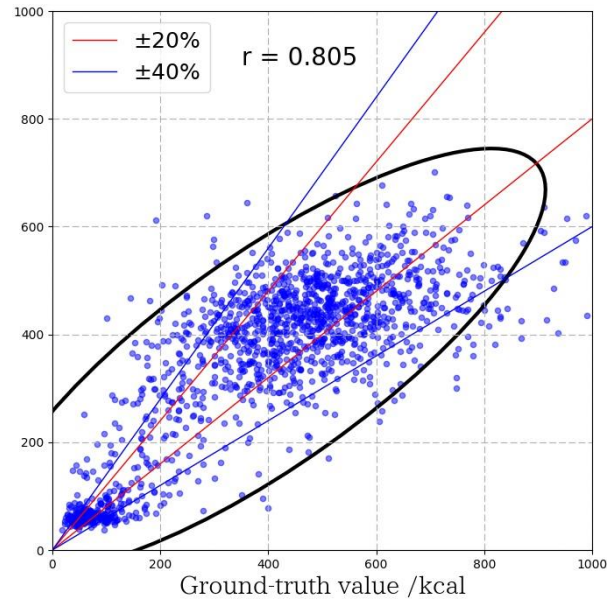
実験結果：手法の比較

推定カロリー値と正解値の相関



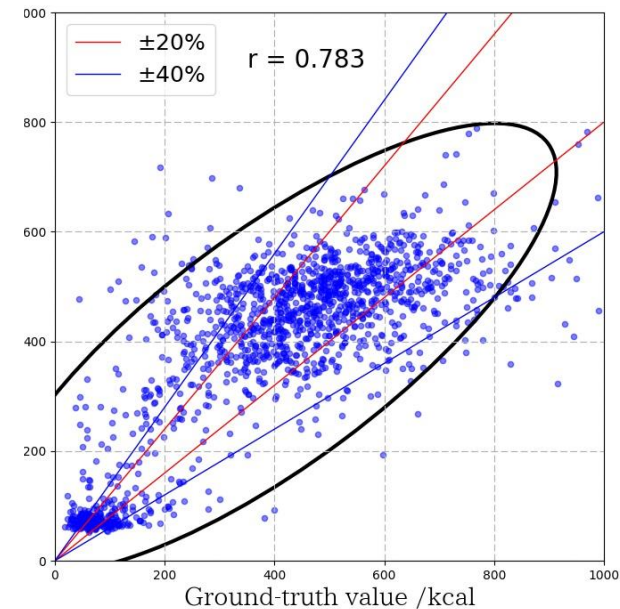
Baseline
(料理カテゴリ分類)

相関値 : 0.784



回帰による手法
(VGG16 multi-task)

相関値 : 0.805

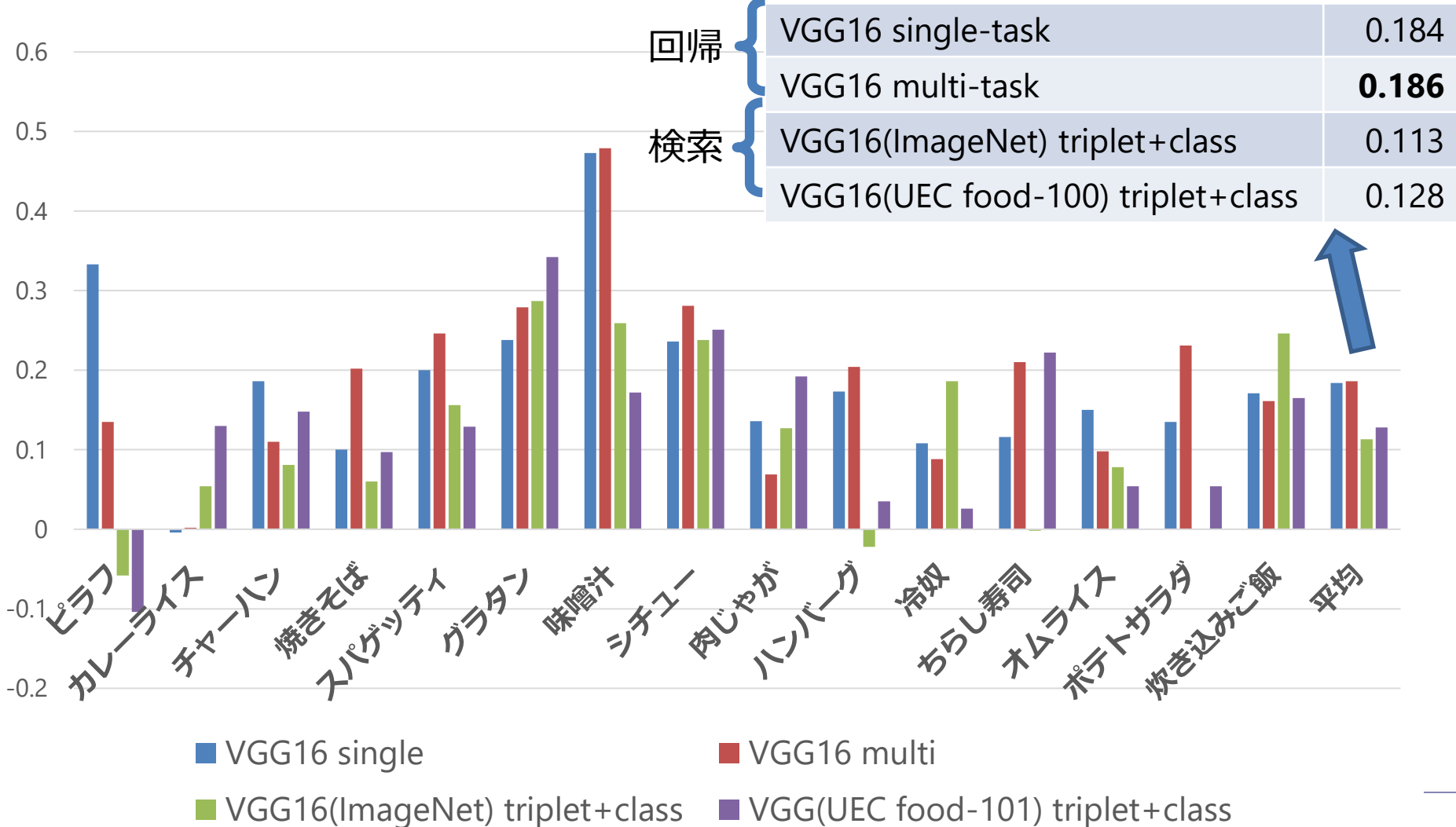


検索による手法
(VGG16(ImageNet)
fc6 triplet + class)

相関値 : 0.783

実験結果：手法の比較

各カテゴリにおける
推定カロリー量値と正解値の相関係数



追加実験: 料理カテゴリを無視した大規模データセットでのカロリー量推定

- 料理 15 カテゴリ以外の画像も使用し, 料理カテゴリを無視した大規模カロリー量付き食事画像データセットを構築. (合計68,774枚)
 - 学習: 80%, テスト: 20%

		相対誤差 (%)	絶対誤差 (kcal)	相関係数	< 20% (相対誤差)(%)
回帰	VGG16 single-task	60.0	132.0	0.436	23.5
	VGG16(ImageNet)	121.5	142.9	0.330	24.7
検索	VGG16(UEC Food-100)	93.9	130.8	0.409	26.8
	VGG16(ImageNet) triplet+class	90.2	128.9	0.428	26.7
	VGG16(UEC Food-100) triplet+class	93.1	131.5	0.396	26.6

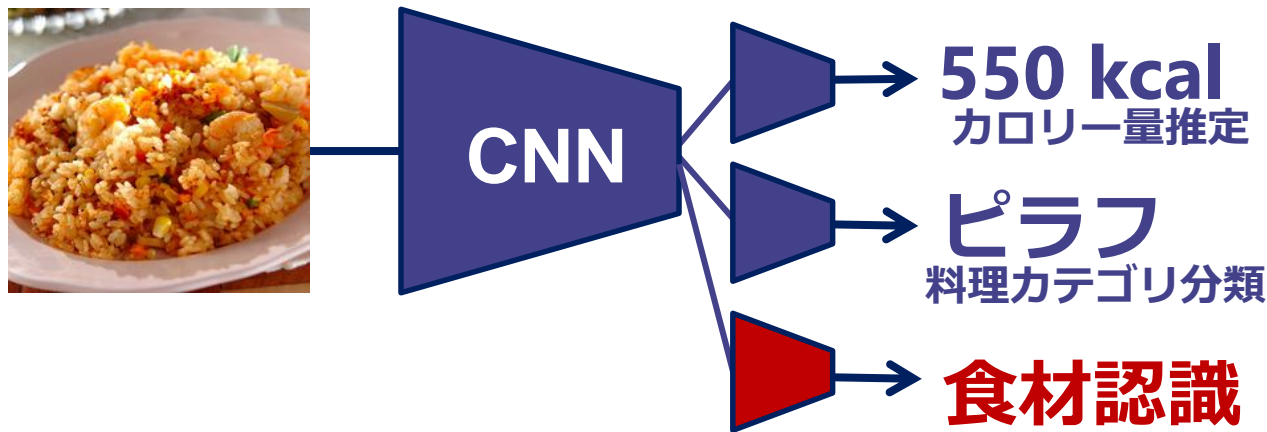
まとめ

- 食事画像からのカロリー量推定について二つの手法の比較を行った。
- 「カテゴリ毎固定」も含めて三つの手法に大きな差は見られなかった。
- 料理カテゴリと同様に食材情報についても考慮する必要があると考えられる。

	平均相関係数	クラス分類(%)
VGG16 single-task	0.184	78.7
VGG16 multi-task (+cat)	0.186	83.2
VGG16 multi-task (+cat+ing+dir)	0.188	85.0

今後

- 回帰: 食材のマルチラベル認識を追加したマルチタスク学習.



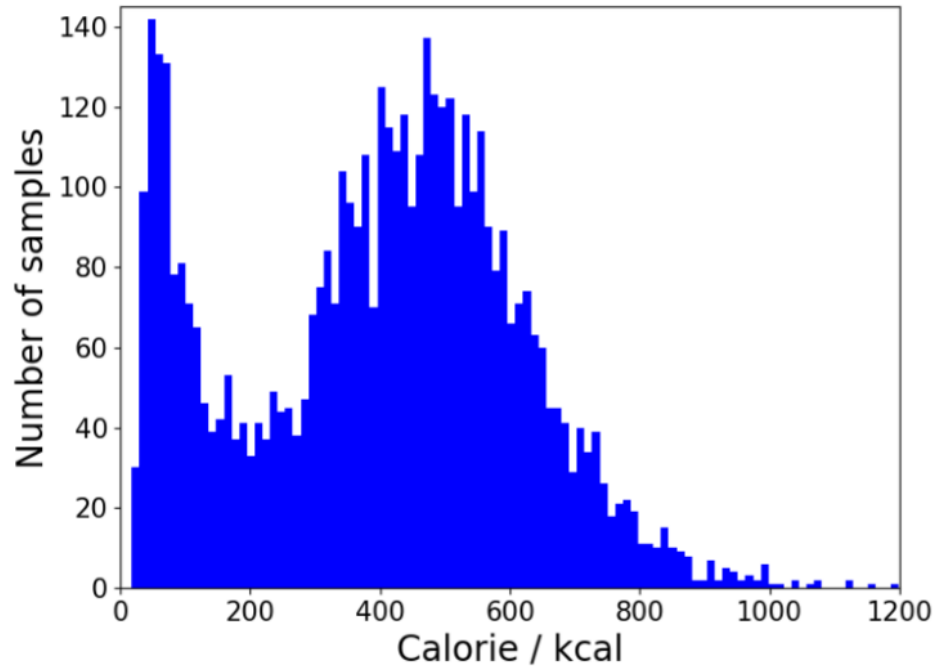
- 検索: 食材情報も含むような画像特徴量の使用.

Salvador, A.ら, "Learning Cross-modal Embeddings for Cooking Recipes and Food Images", CVPR, 2017

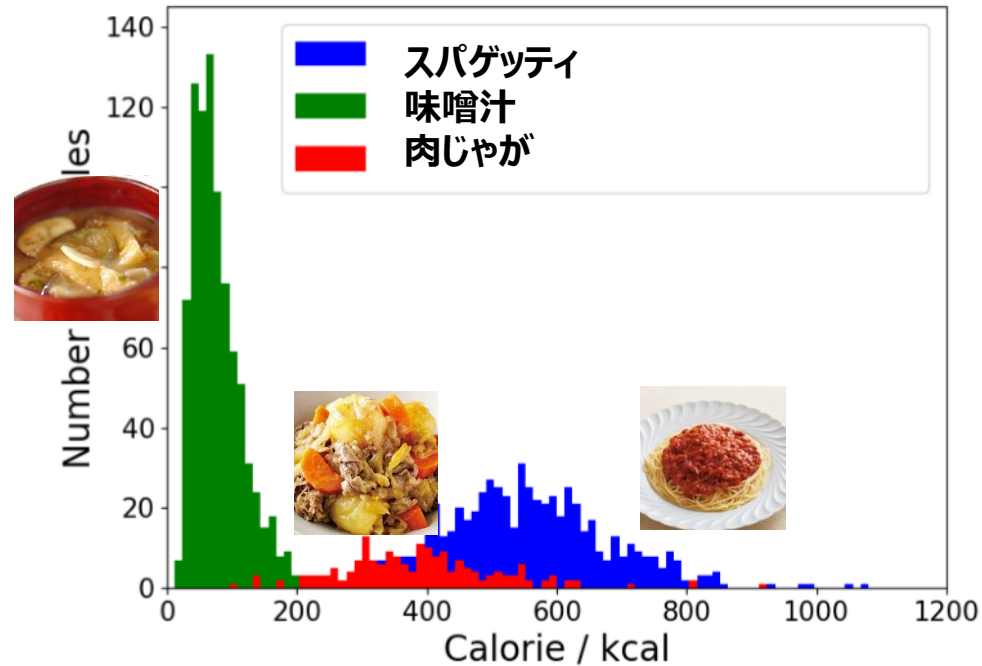


カロリー量付き食事画像データセット

データセットのカロリー量分布



全体



スパゲッティ, 味噌汁, 肉じゃが.

カロリーー量推定損失関数

- 絶対誤差, 相対誤差

	rel. err. (%)	abs. err. (kcal)	correlation	20% err. (%) √I	40% err. (%) √I
only L_{re}	29.4	105.9	0.776	42.0	77.0
only L_{ab}	59.8	134.9	0.589	36.7	63.5
$L_{re} + L_{ab}$	29.4	100.7	0.778	45.9	77.7

実験結果：手法の比較

- VGG16を用いた場合の比較.
- テストにはカロリー量付き食事画像データセットを使用.
- Baseline (クラス分類) : 料理カテゴリごとにカロリー値を固定.

		相対 誤差 (%)	絶対 誤差 (kcal)	相関 係数	< 20% (相対誤 差)(%)
固定	Baseline	32.4	93.6	0.784	50.0
回帰	VGG16 single-task	29.2	100.4	0.783	46.0
	VGG16 multi-task	28.0	96.5	0.805	47.2
*参考	*VGG16 multi-task(+ 食材, 調理手順)	27.4	91.2	0.817	50.1
検索	VGG16(UEC food-100)	38.0	101.0	0.764	47.2
	VGG16(ImageNet) triplet + class	34.0	94.5	0.783	49.5
	VGG16(UEC food-100) triplet + class	36.7	98.6	0.777	48.1