

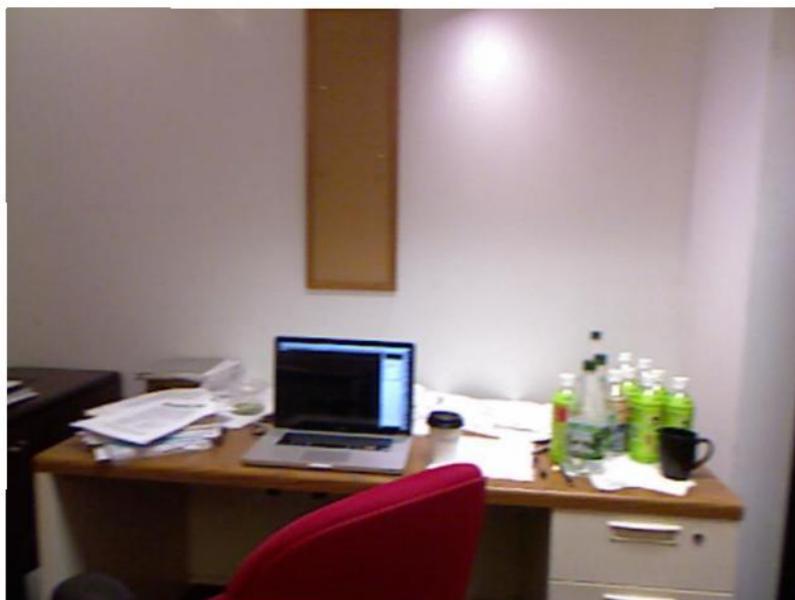
# 物体の前後関係情報を使った 単眼深度推定

高嶺みちる

# 深度推定

(奥行き情報)

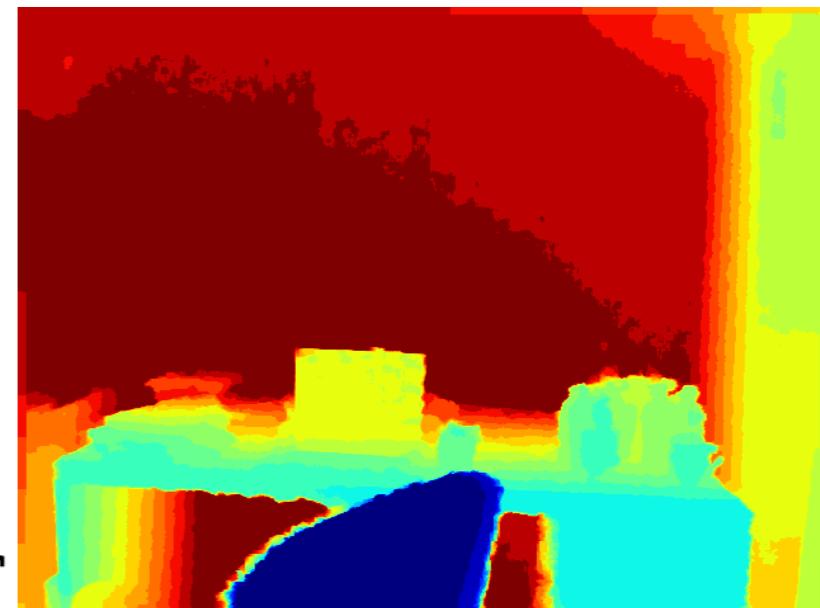
RGB画像



Far

Near

RGB-D画像

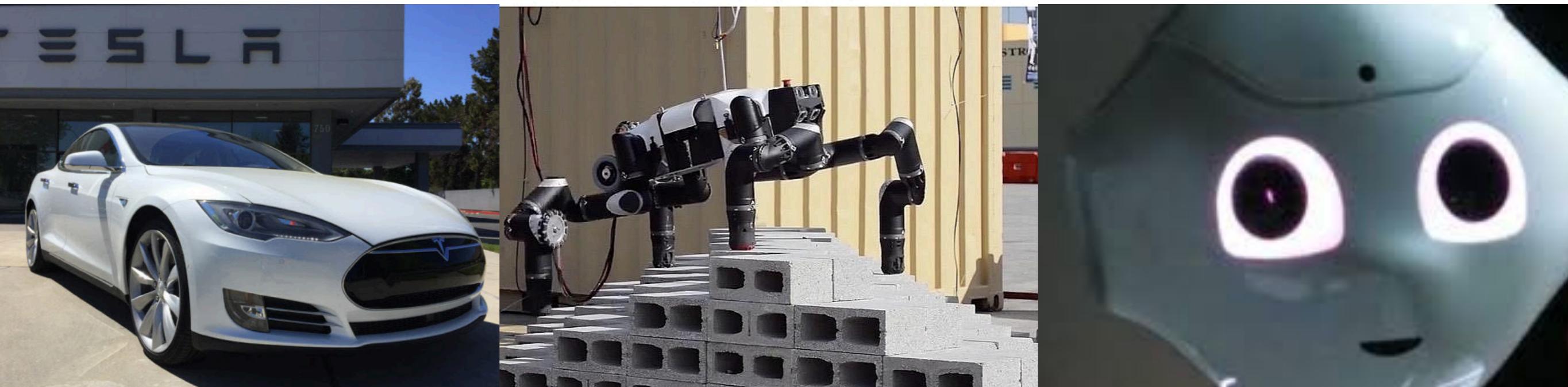


カメラから被写体までの距離を取得する

概要 先行研究 提案手法 実験 考察

## 深度推定

自動運転 災害時救助ロボット ペッパーくん



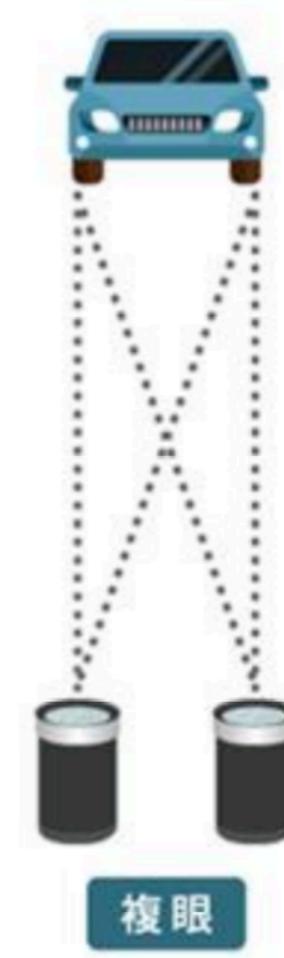
機械が空間情報を取得するための重要な手段

# 单眼深度推定

单眼=

RGB画像1枚

を使った深度推定



# なぜ単眼深度推定なのか

単眼=



## メリット

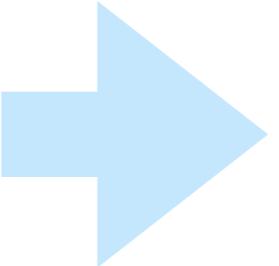
- ・センサーが安い
- ・汎用的な形式
- ・省スペース

## デメリット

- ・困難さ
- ・原理的に不可能
- ・経験則がでかい

# 問題領域の分割1

この写真は何を表している？



問1 特徴的な模様は？

問2 人間はあるか？

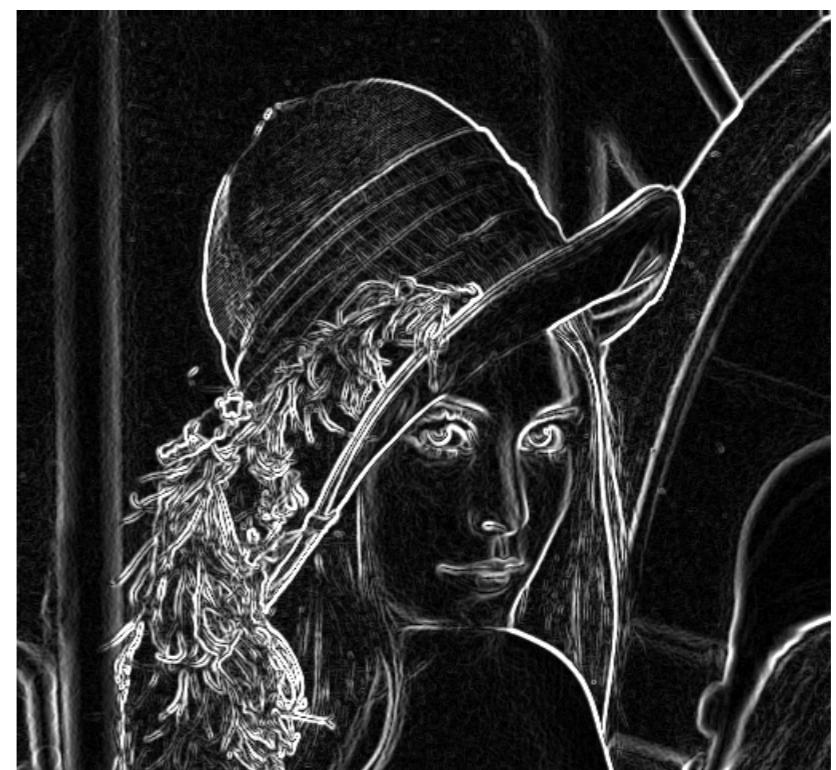
問題領域の分割

1つの問題を解くためにより簡単な問題に分ける

## 現在のアプローチ



- ・ カテゴリ分類  
精度向上したが**深度が離散**  
**曖昧なラベルを扱えない**



- ・ 深度勾配推定  
精度向上  
推定結果の外れ値削減

## 分割方法の明暗



数値として定義できない

数値として定義できる

数値として定義できるか否かが分かれ目？  
条件を満たすオブジェクト情報を考える

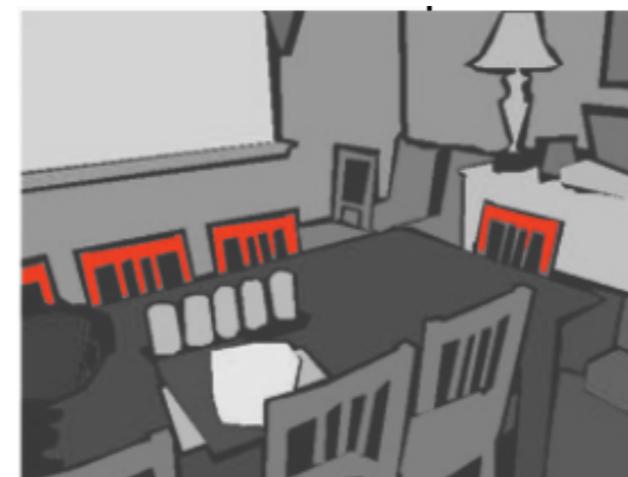
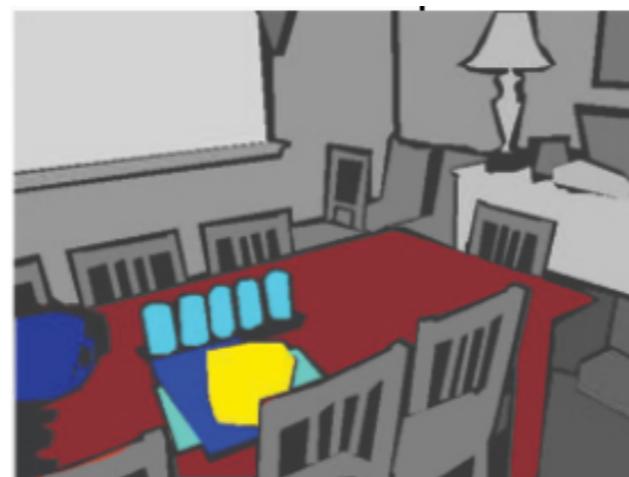
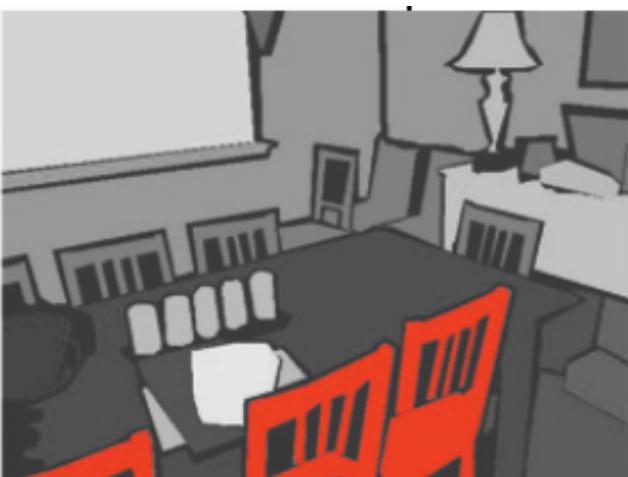
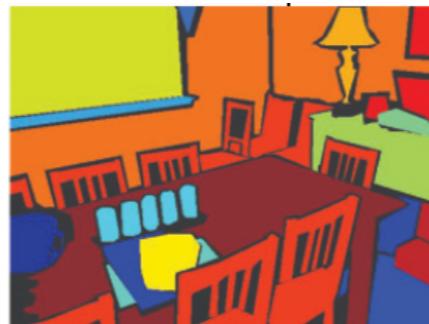
## 背景と目的

- ・人間は問題領域を分割して高度な深度推定を実現
- ・单眼深度推定においても場合によっては問題領域の分割が有効
- ・数値として定義可能か否かが適性の分かれ目と推測

- ・深度推定問題を重なり情報推定問題に分割することで……
- ・单眼深度推定の精度向上
- ・扱える画像の種類を増やす

# コンセプト

大まかな前後関係情報の推定と  
深度の推定に問題領域を分割



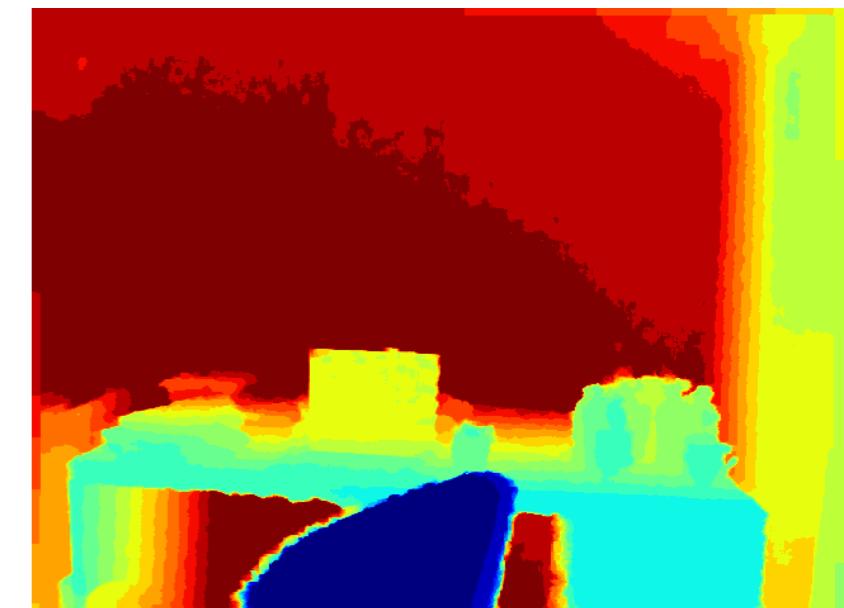
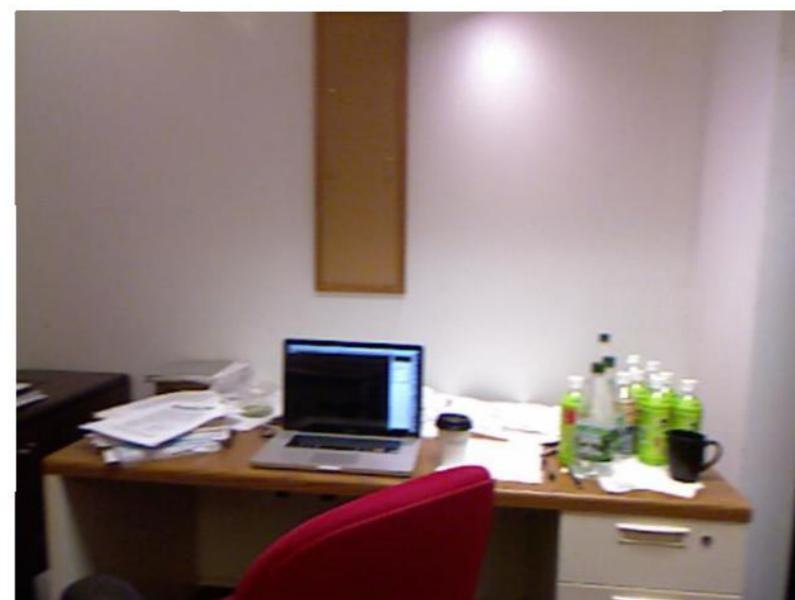
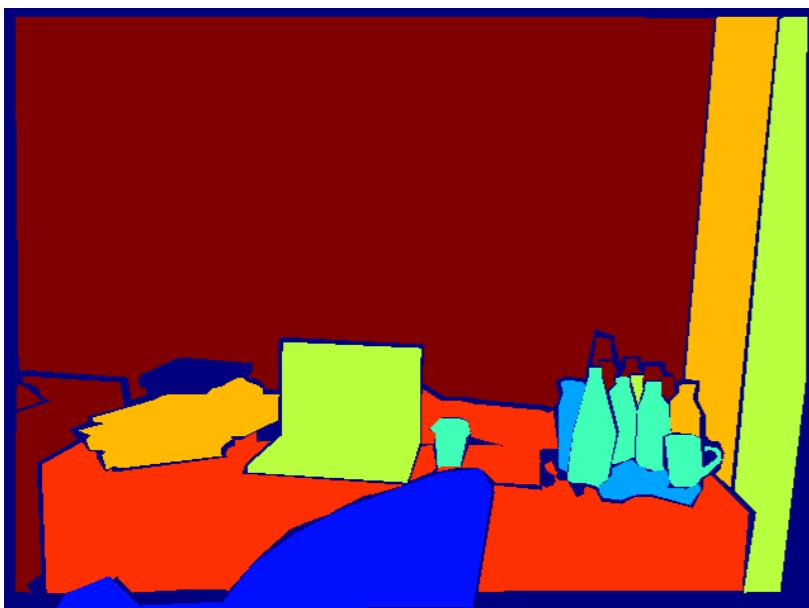
①層目

②層目

③層目

④層目

## 深度との違い



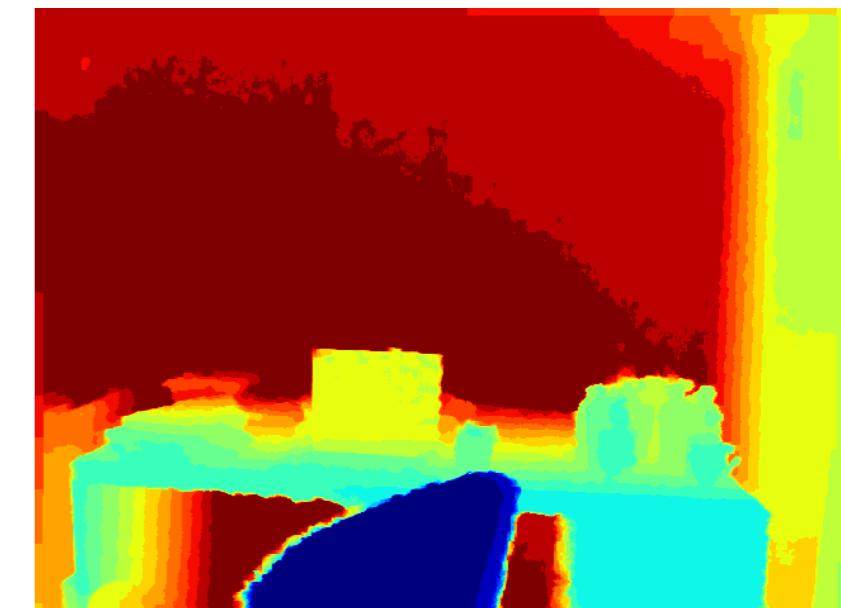
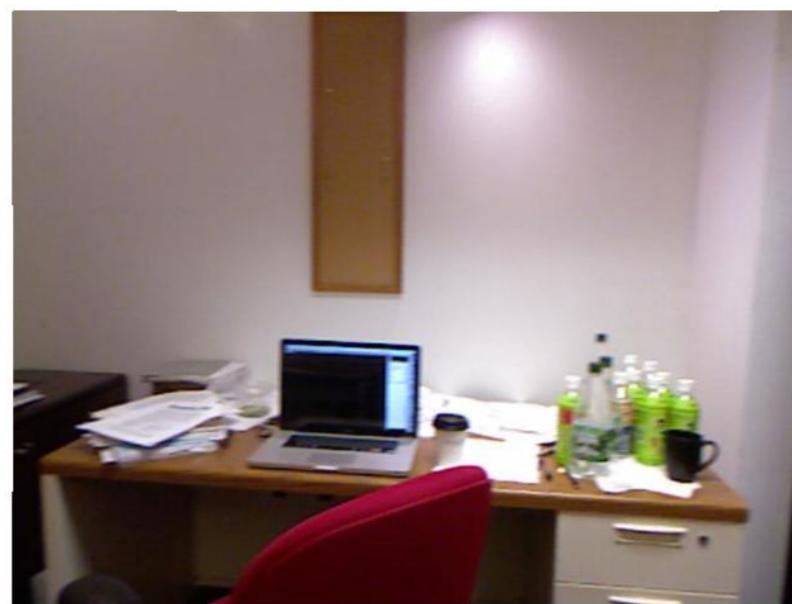
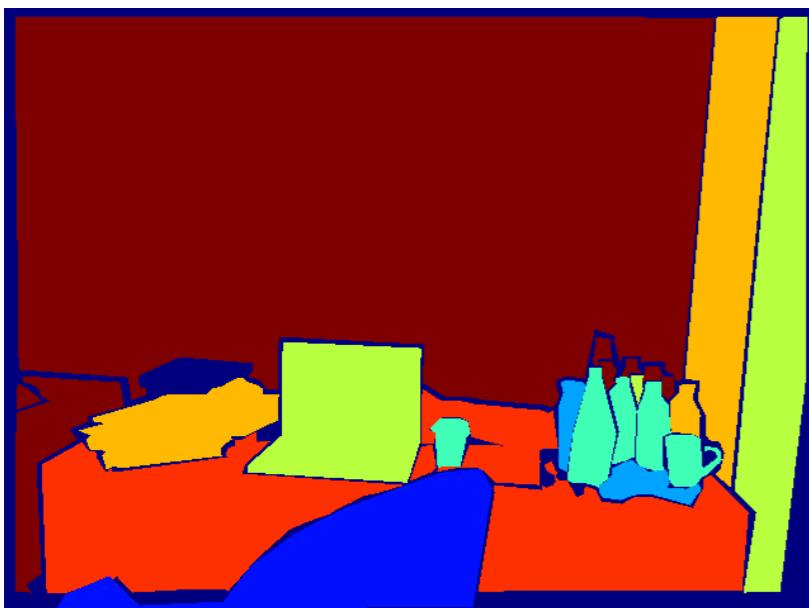
重なり

- 分類問題
- [0~N]の整数值

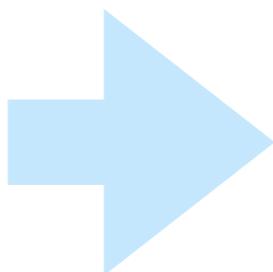
深度

- 回帰問題
- [0~ $\infty$ ]の連続値

# 重なり情報を利用する意味

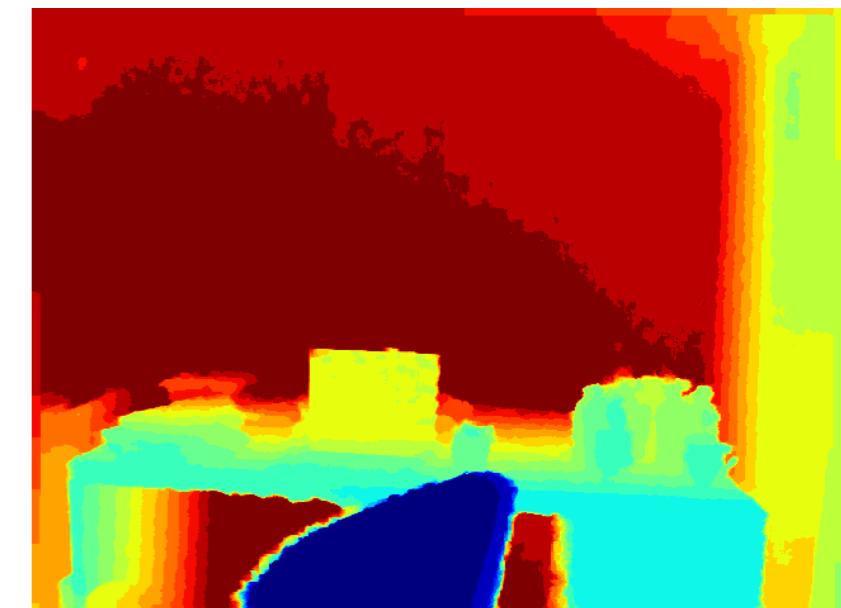
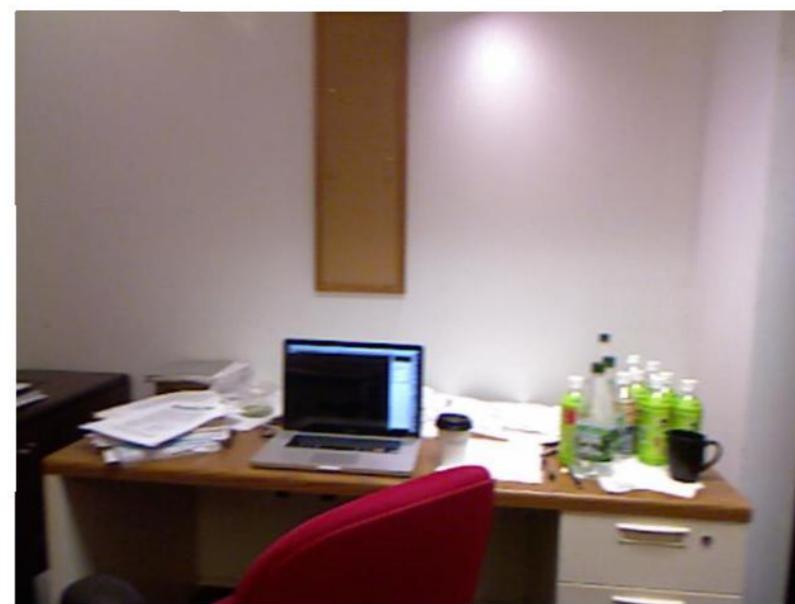
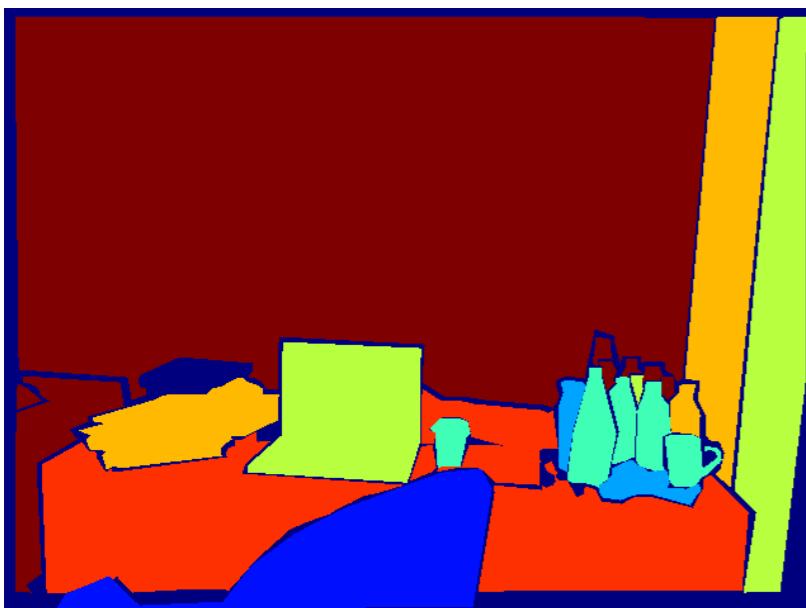


深度よりも楽な  
重なり情報を  
ヒントとして推定



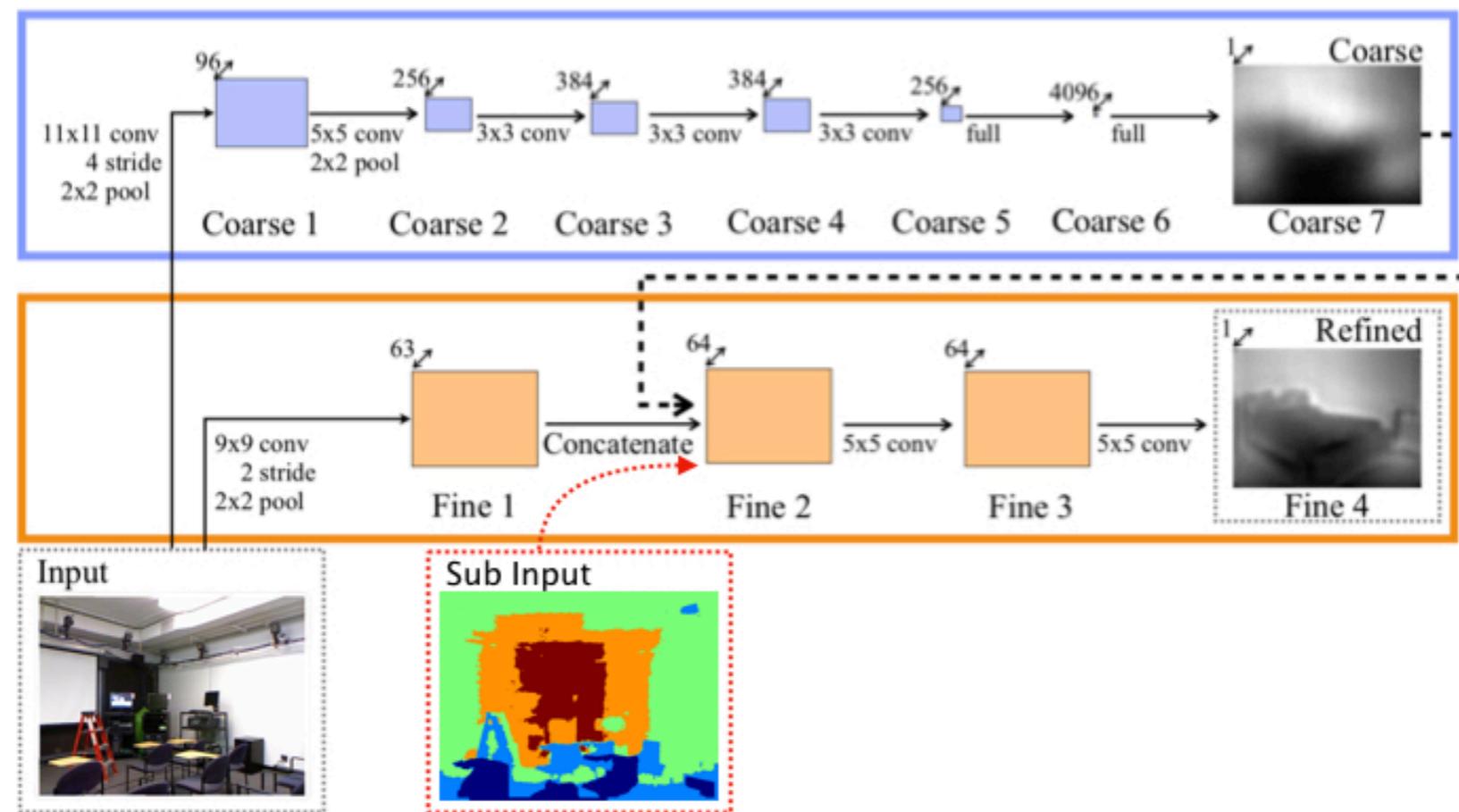
ヒントを元に  
深度を推定

# 重なり情報の利点



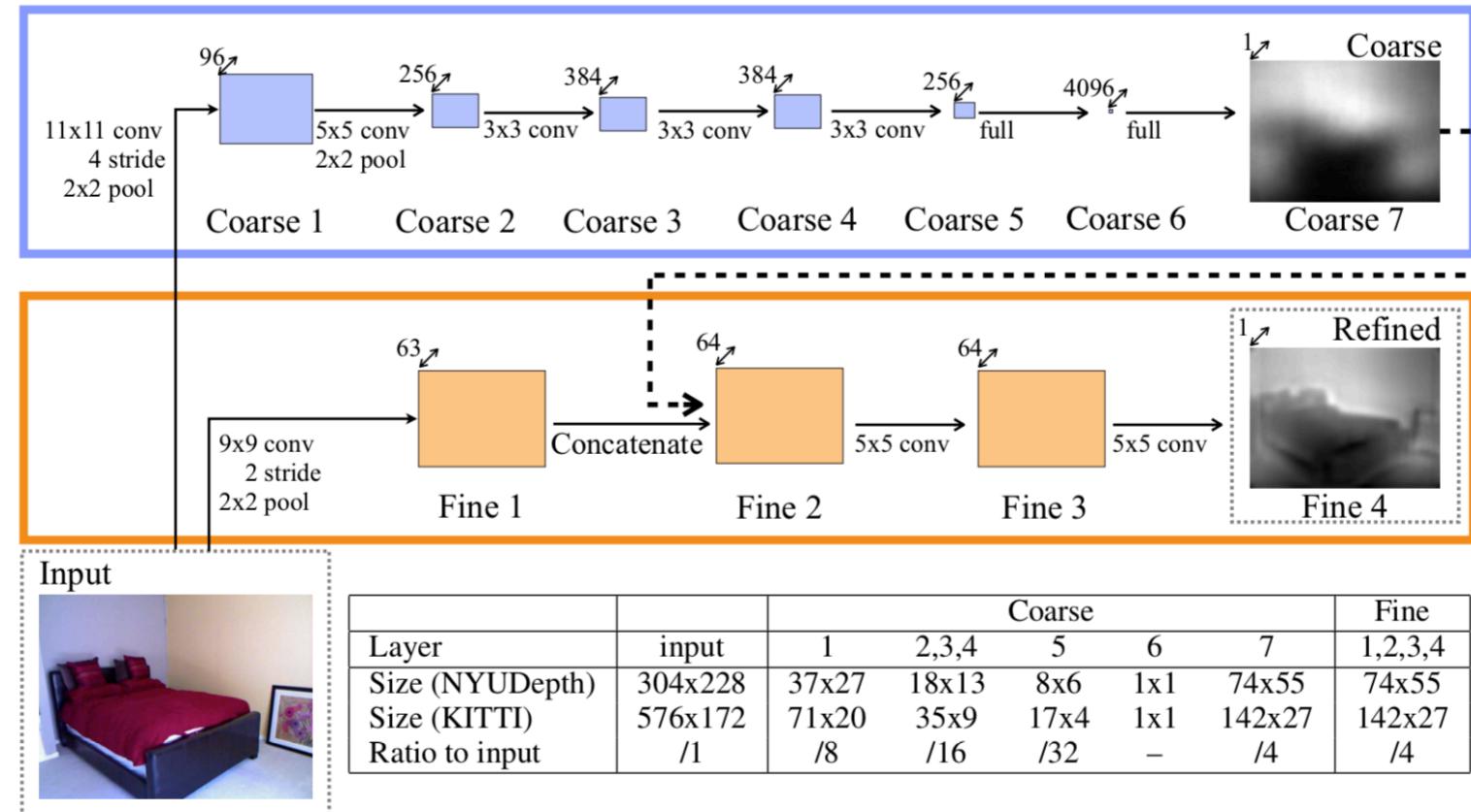
- ・重なり情報を求めるることは  
**深度推定を分類問題に落とし込むことと同義**
- ・重なり情報を介して深度推定を行うことで  
**分類問題へのアプローチを深度推定に流用できる**

# 重なり情報の活用



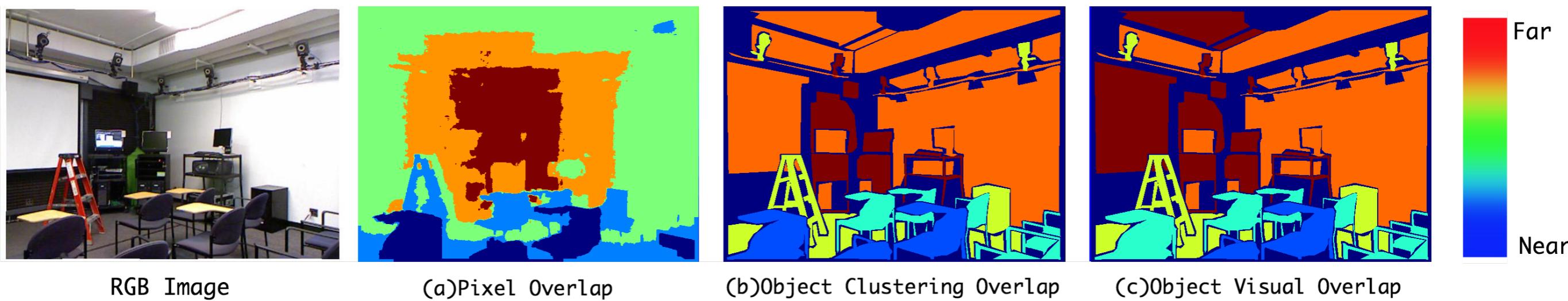
- ・深度を求める前段階としてRGB画像から重なり画像を生成する
- ・生成した重なり画像をMSMにおける大域情報として扱い  
推定モデルの中間層に入力として与え、深度推定の向上を図る

# EigenらのMSM



- CNN で人間の深度推定を模倣
- 大域情報と局所情報を段階的に推定
- 2つの情報を分けて考える利点は問題領域の分割

# 重なり情報の定義



RGB Image

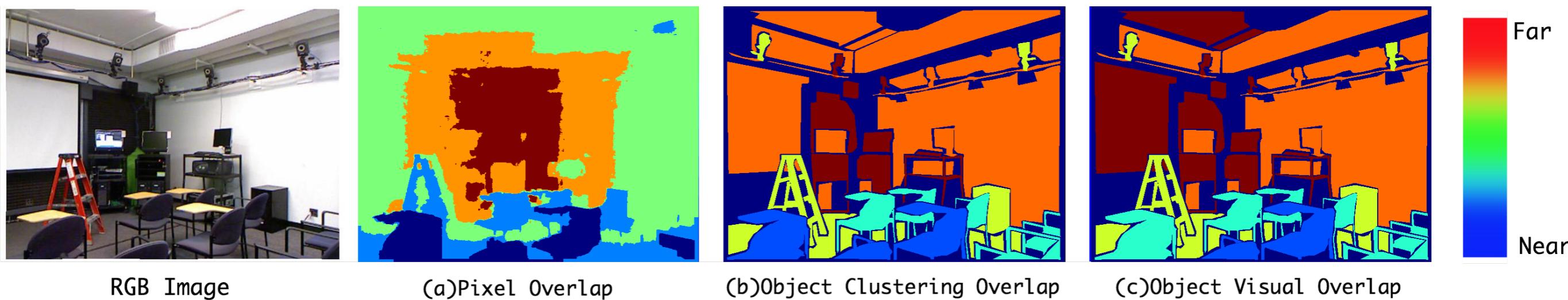
(a)Pixel Overlap

(b)Object Clustering Overlap

(c)Object Visual Overlap

- ・重なり情報 物体もしくは空間の前後関係の順番のみに着目した情報
- ・重なり画像 カメラに最も近い領域には0、次に近い領域には1以下昇順でpixelごとに整数値の番号が割り振られた画像

# 重なり情報の形態



## 1. Pixel Overlap

pixelを深度でクラスタリングしクラスタ毎に番号を振る

## 2. Object Clustering Overlap

属している深度の区域に応じてオブジェクトに番号を振る

## 3. Object Visual Overlap

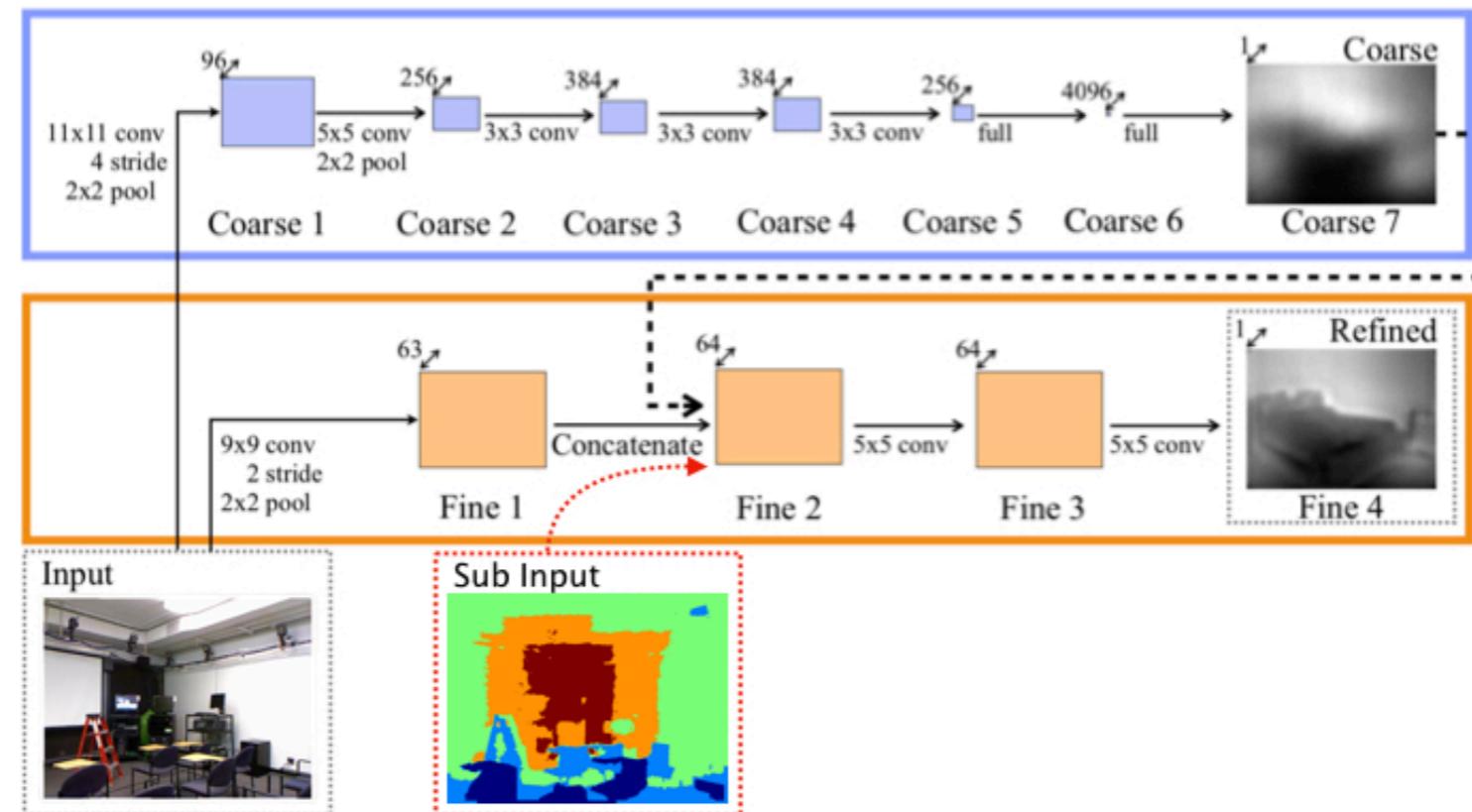
遮った画面上の物体の数に応じてオブジェクトに番号を振る

概要 先行研究 提案手法 実験 考察

## 実験の概要

重なり情報が与えられた場合  
深度推定に役立つか？

# 実験の概要



・重なり情報が深度推定の精度向上に寄与するという仮説を検証

1. MSMに重なり画像を**与えずに**学習させる

a. RGB画像のみを入力 b. RGB画像+ラベル情報を入力

2. MSMに重なり画像を**与え**学習させる

## 比較

- ・ 2種類の評価関数を用いた比較
- ・ ラベル情報への依存度の比較

# 重なり情報なしとの比較

	Plane	Pixel	Cluster	Visual
$\delta < 1.25$	0.341	<b>0.465</b>	0.396	0.424
$\delta < 1.25^2$	0.601	<b>0.783</b>	0.682	0.724
$\delta < 1.25^3$	0.764	<b>0.927</b>	0.840	0.881
MSE	0.0258	<b>0.0112</b>	0.0189	0.0145

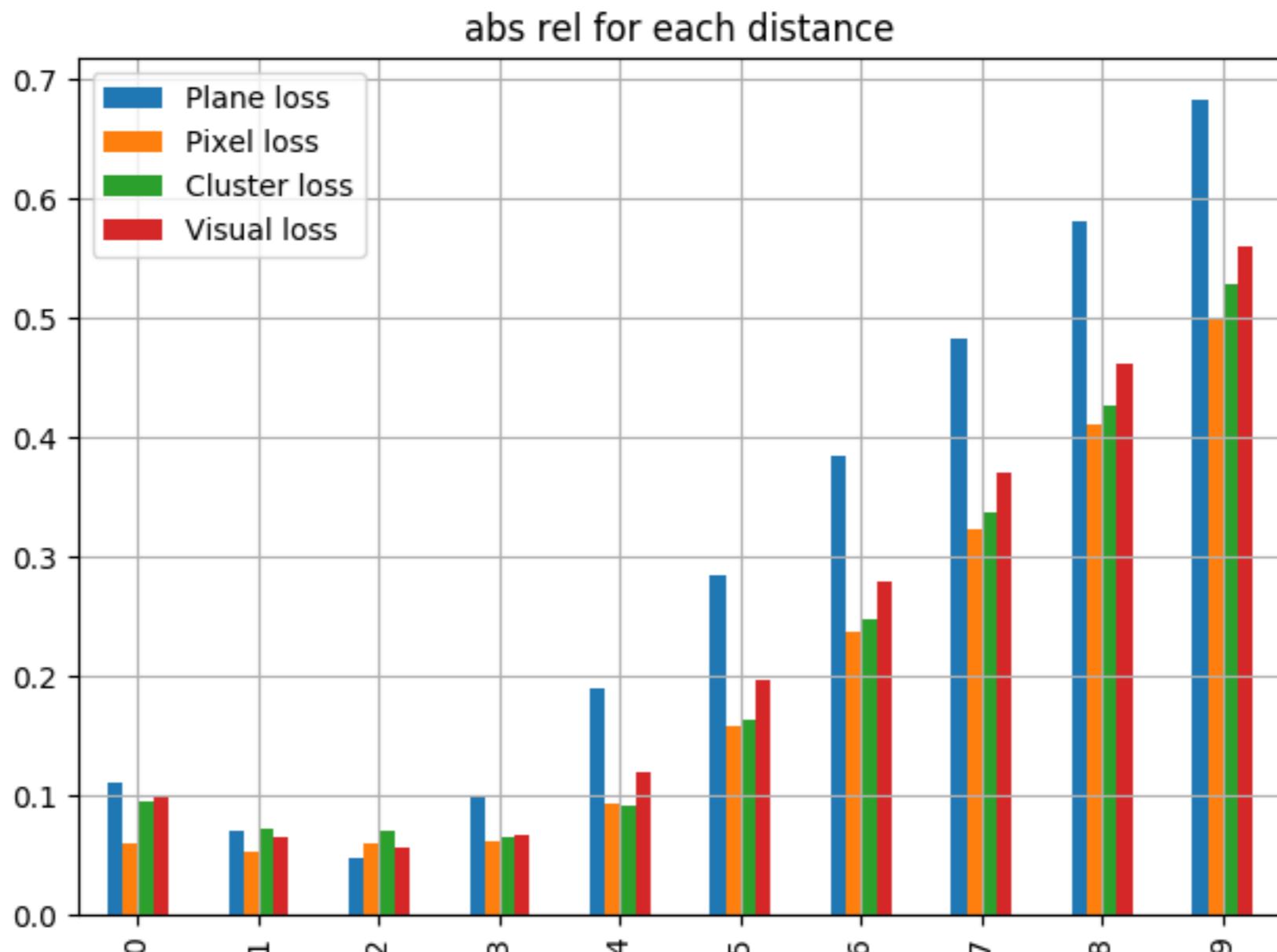
Planeモデルよりも  
優れた推定精度を記録

## ラベル情報ありとの比較

	Semantic	Pixel	Cluster	Visual
$\delta < 1.25$	0.434	<b>0.465</b>	0.396	0.424
$\delta < 1.25^2$	0.737	<b>0.783</b>	0.682	0.724
$\delta < 1.25^3$	0.889	<b>0.927</b>	0.840	0.881
MSE	0.0138	<b>0.0112</b>	0.0189	0.0145

Pixelモデルが  
Semanticモデルより優れた

## 評価関数による比較



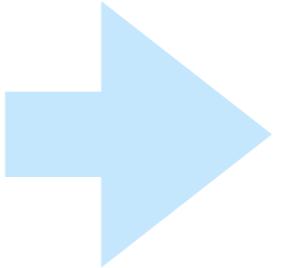
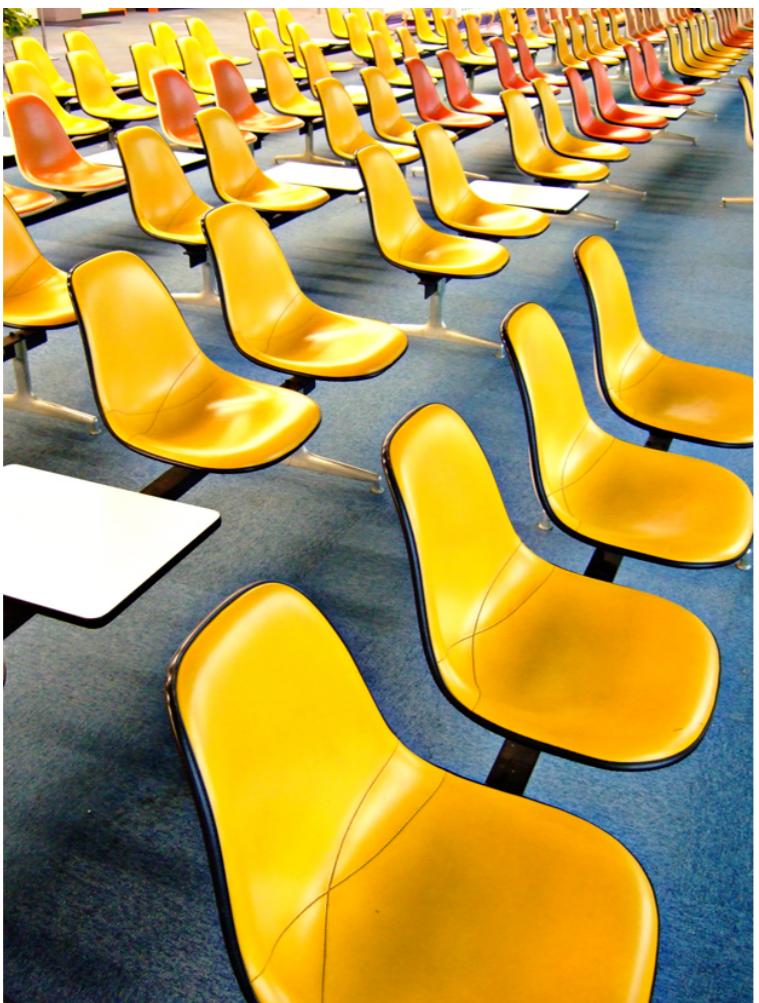
0~10mを1m区間で区切った際のloss

青が従来手法

## 比較

- ・ 2種類の評価関数を用いた比較
- ・ ラベル情報への依存度の比較

## ラベル情報への依存性

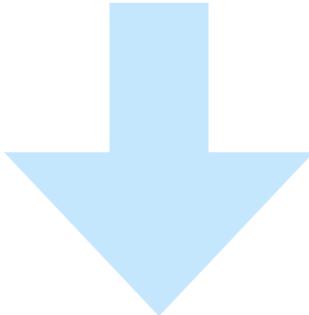


椅子だから2m!

- ・ カテゴリ分類に分割した際に想定される問題  
RGB画像に関わらずラベルの平均深度を出してしまう

## ラベル情報への依存度

平均深度を常に出してしまう



各ラベル領域の深度の分散が小さくなる

# ラベル情報への依存度



奥の椅子は10mで  
手前は1mですね

正解深度に対する標準偏差A

うーん、  
全部2m！

予測深度に対する標準偏差B

推定精度が低いほど  
分散の値が低いほど

$$\sum_{n=1}^N \frac{|A_n - B_n|}{B_n}$$

依存度が高くなる

## 依存度による比較

Plane	Semantic	Pixel	Cluster	Visual
9.485	6.482	1.846	1.7734	3.064

重なり情報3種類とともに

Planeモデル

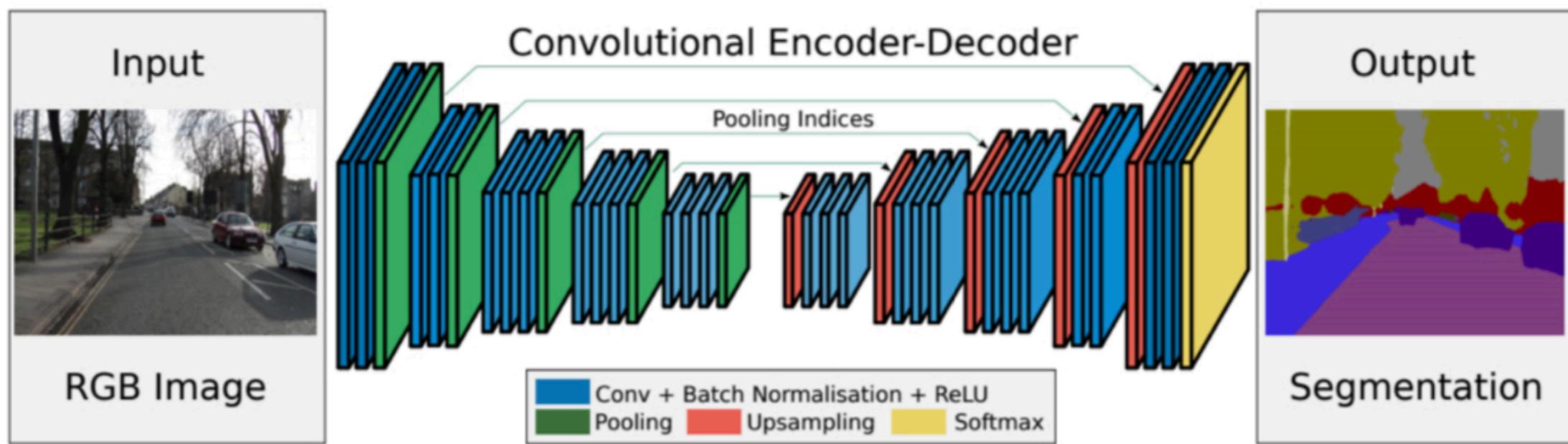
Semanticモデルよりも依存度が低い

概要 先行研究 提案手法 実験 考察

## 実験の概要

重なり情報を  
実際に推定することができるのか？

## 実験の概要



- 重なり情報の推定可能性を検証する

- 1.SegNetを重なり画像を推定するように学習させる
- 2.推定した重なり画像をMSMに与えて学習させる

## 比較

- ・ 2種類の評価関数を用いた比較
- ・ ラベル情報への依存度の比較

## 評価関数による比較

	Plane	Get Pixel	Get Cluster	Get Visual
$\delta < 1.25$	0.341	<b>0.490</b>	0.403	0.434
$\delta < 1.25^2$	0.601	<b>0.787</b>	0.679	0.735
$\delta < 1.25^3$	0.764	<b>0.920</b>	0.823	0.892
MSE	0.0258	<b>0.0120</b>	0.0197	0.0142

重なり情報3種類とともに  
Planeよりも優れた評価値を記録

# 評価関数による比較

	Semantic	Get Pixel	Get Cluster	Get Visual
$\delta < 1.25$	0.434	<b>0.490</b>	0.403	0.434
$\delta < 1.25^2$	0.737	<b>0.787</b>	0.679	0.735
$\delta < 1.25^3$	0.889	<b>0.920</b>	0.823	0.892
MSE	0.0138	<b>0.0120</b>	0.0197	0.0142

Pixelモデルが  
Semanticモデルより優れた

## 比較

- ・ 2種類の評価関数を用いた比較
- ・ ラベル情報への依存度の比較

# 依存度による比較

Plane	Semantic	Get Pixel	Get Cluster	Get Visual
9.485	6.482	3.005	4.050	5.504

重なり情報3種類とともに

Planeモデル

Semanticモデルよりも依存度が低い

## まとめ

单眼深度推定を重なり情報推定問題に分割することで

- ・予測精度の向上
- ・ラベル情報への依存度の低下

重なり情報の推定可能性については問題が残る

- ・現時点では最大でも45%前後の推定精度

	Get Pixel	Get Cluster	Get Visual
accuracy	0.402	0.440	0.463

## 考察

- ・オブジェクト領域毎に深度の類似度を推定することで  
深度推定モデルのラベル依存度を大きく改善できる

Plane	Semantic	Pixel	Cluster	Visual
9.485	6.482	1.846	1.7734	3.064

- ・現在は、最も通常の深度に近いPixelモデルが有利  
推定のしやすさでVisualモデル優っている

	Get Pixel	Get Cluster	Get Visual
accuracy	0.402	0.440	0.463