

B51. 単眼深度推定のアンサンブル

早稲田大学大学院 先進理工学研究科 電気・情報生命専攻 村田研究室
5326E086 吉田 雄紀 Yuki Yoshida

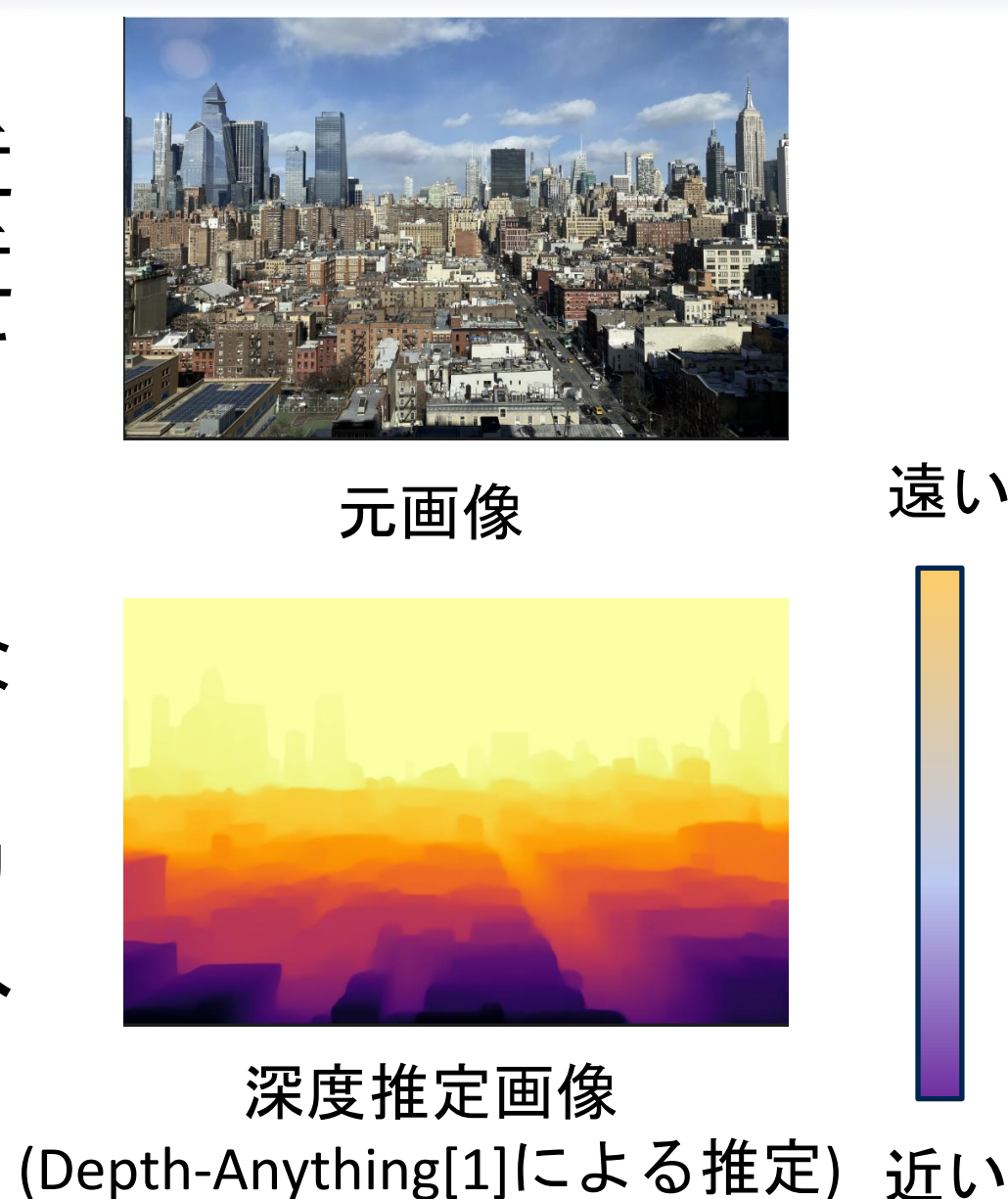
0. 概要

本研究では単眼深度推定に対して複数のモデルの出力分布を統合するアンサンブル手法を検討する。単眼深度推定では、モデルごとに推定傾向が異なるため、複数モデルを統合することで、より安定した推定が期待できる。特にDiffusion Modelを用いた深度推定モデルでは初期ノイズを変更することで、複数の深度推定結果を得られるため、推定結果を確率分布として扱える。本研究では複数モデルの出力分布を統合し、推定精度だけでなく、信頼度区間も評価することを目指す。

1. 背景：単眼深度推定とは

・単眼深度推定とは、一枚の画像からどこが近くてどこが遠いのかを推定する技術である。近年では深層学習を用いる手法が広く研究されている。

・単眼深度推定は、自動運転やロボティクスなどへの応用が期待されている。また、LiDARなどの距離センサーを用いる方法と比較して、カメラ画像のみで推定できるため低コストに導入でき、現在盛んに研究が行われている。



2. 問題設定

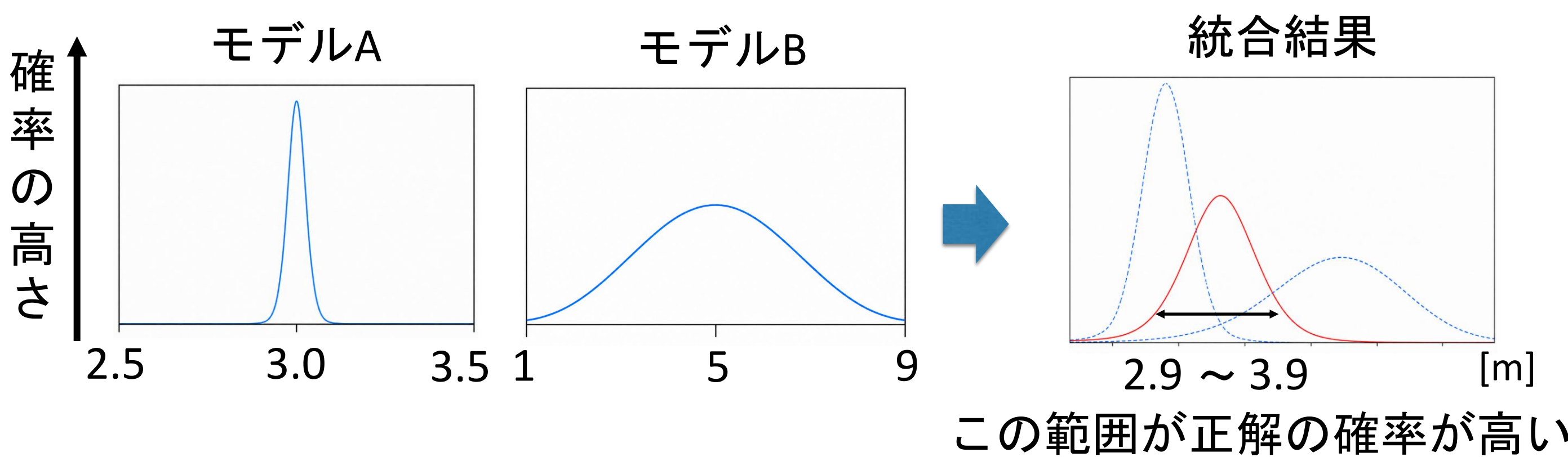
・単眼深度推定のアンサンブル
単眼深度推定モデルでは各モデルごとに得意不得意が違う事がある。例えばあるモデルは「近い物体が得意」だったりあるモデルは「遠い領域が得意」のような得意不得意が異なる。そこで複数のモデルを組み合わせることで、お互いの弱点を補完する推定を目的とする。

・統合方法について
本研究では各深度マップについて「この距離である可能性がどれくらい高いか」という確率分布を考え、その確率分布同士を統合する。分布で考えることで、推定深度だけでなく、もっともらしい深度の範囲を評価できる。

・単純な平均の場合



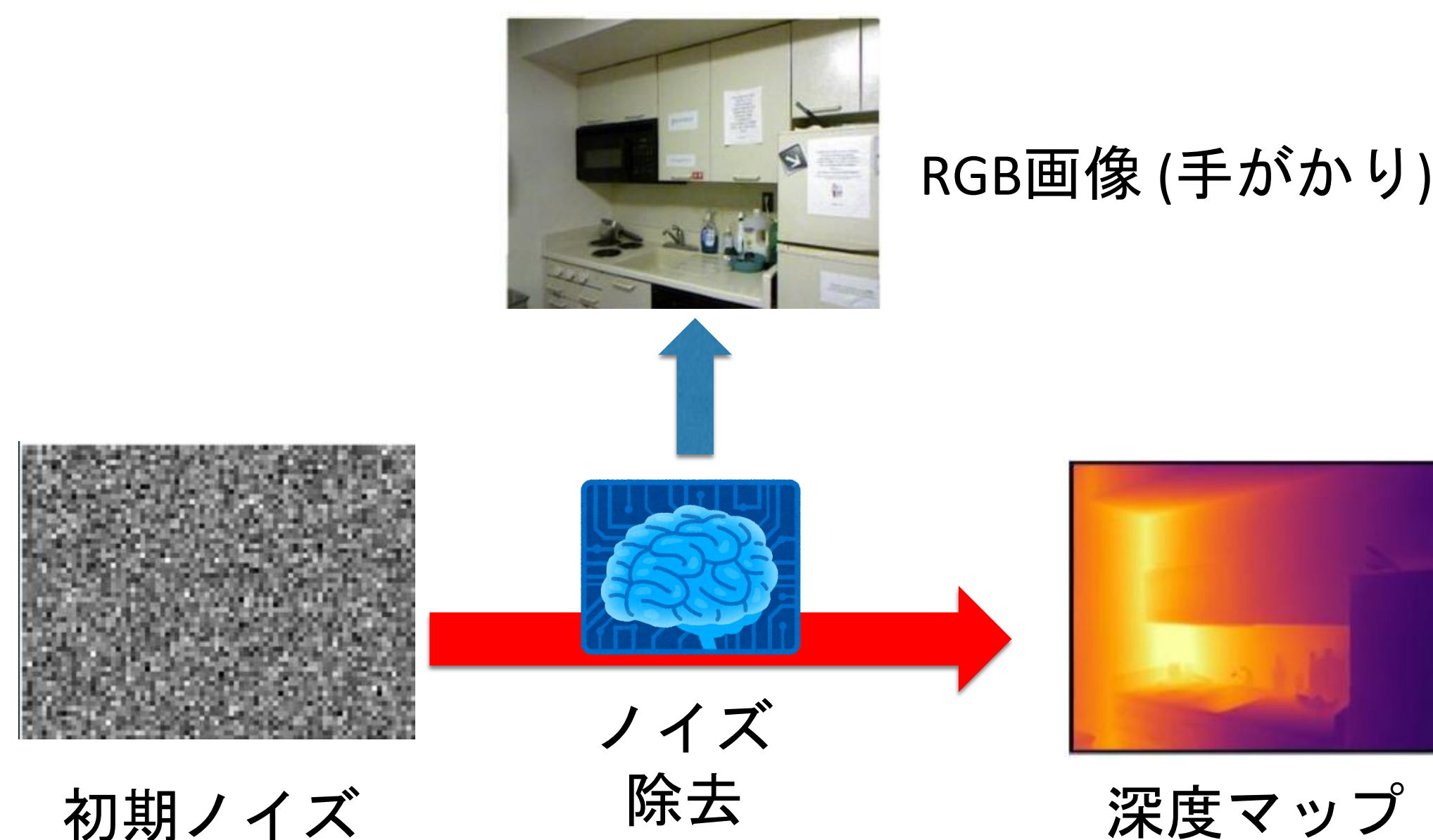
・分布統合の場合



3. 先行研究

・Diffusionを用いた単眼深度推定モデル

Diffusionベースの単眼深度推定モデル[2]では、RGB画像を手がかりとして、任意の初期ノイズからあり得そうな深度マップを生成する。

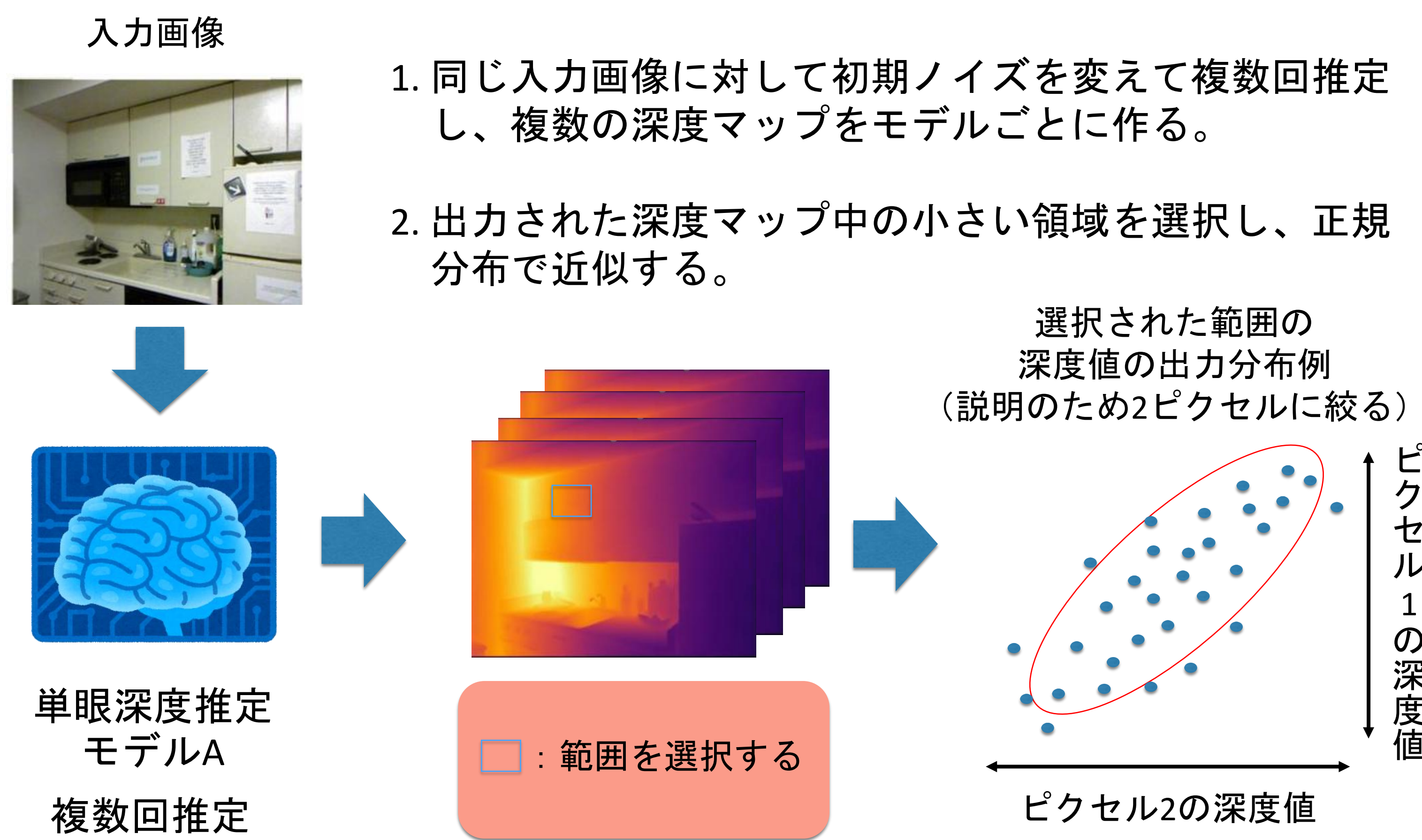


・なぜDiffusionを用いた単眼深度推定モデルを扱うのか

Diffusionベースの単眼深度推定モデルでは初期ノイズを変更することで、同じ画像から複数の深度マップを得ることができる。複数の深度マップを扱う事で、「この距離である可能性がどれくらい高いか」という分布を作ることができる。

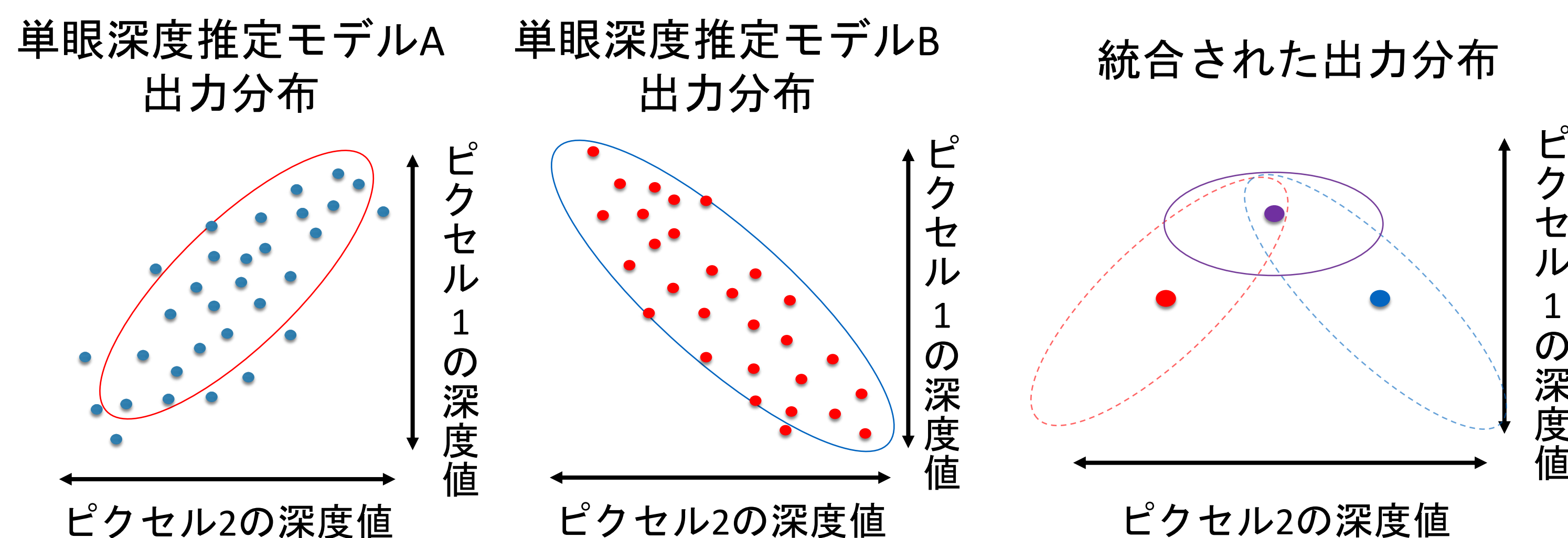
This research investigates an ensemble method for monocular depth estimation by integrating the output distributions of multiple models. In monocular depth estimation, each model may have different estimation tendencies; therefore, integrating multiple models is expected to enable more stable estimation. In particular, diffusion-based depth estimation models can generate multiple depth estimation results by changing the initial noise, allowing the estimation results to be treated as probability distributions. This research aims to integrate the output distributions of multiple models and evaluate not only estimation accuracy but also confidence intervals.

4. 手法



3. 以上のような操作を複数のモデルで行うことで、各モデルの出力分布を得ることができる。

4. 得られた各モデルの出力分布を複数モデルが共通して支持する深度を強調できるe-mixture（分布の積による統合）を用いて分布同士を統合する。



5. 統合された出力分布について、もっともらしい深度値の範囲と、分布の平均値を用いた推定精度を評価する。

5. 今後の展望

・各分布に割り当てる重みの検討
分布統合では、各モデルの出力分布をどの程度反映するかを決める重みが重要となる。今後は、入力画像や画像領域に応じて各モデルの重みを決定するゲートモデルを導入し、MoE (Mixture of Experts) [3]に基づく重み付き分布統合を検討する。

・バリエーションを持った単眼深度推定モデルの学習
アンサンブルによって精度を上げるためには、各モデルが同じような誤りをしないことが重要である。そこで学習データを変更したモデルや、一部の構造を変更したモデルを用意することで、異なる特徴を持つモデルを複数作成する。

6. 参考文献

[1]. Yang et al., "Depth Anything," CVPR, 2024.
[2]. Ke et al., "Repurposing Diffusion-Based Image Generators for Monocular Depth Estimation," CVPR, 2024
[3]. Jacobs et al., "Adaptive Mixtures of Local Experts," Neural Computation, 1991.