

姿勢推定の基本と ディープラーニングを用いた実例

AI演習

<https://www.kkaneko.jp/ai/ae/index.html>

金子邦彦



- 
- ①姿勢推定の基本技術**
 - ②姿勢推定のバリエーションと実応用**
 - ③実用スキルの向上**

アウトライン

1. イントロダクション
2. 姿勢推定の基本
3. 人体の姿勢推定
4. ディープラーニングによる姿勢推定の仕組み
5. 頭部の姿勢推定、物体の姿勢推定

Google Colaboratory



Colaboratory へようこそ

```
x = [5, 4, 1, 3, 2]
for i in x:
    print(i * 120)
```

600
480
120
360
240

[1] x = 100

```
if (x > 20):
    print("big")
else:
    print("small")
```

big

s = 0
for i in [1, 2, 3, 4, 5]:
 s = s + i
print(s)

15

URL: <https://colab.research.google.com/>

- ・オンラインで動く
- ・Python のノートブックの機能を持つ
- ・Python や種々の機能がインストール済み
- ・本格的な利用には、Google アカウントが必要

Google Colaboratory の全体画面



メニュー
(目次, 検索と置換,
変数, ファイル)

The screenshot shows the Google Colaboratory interface. At the top is a toolbar with navigation icons (back, forward, search, etc.). Below it is a menu bar with 'Colab' (PRO), 'ファイル', '編集', '表示', '挿入', 'ランタイム', 'ツール', and 'ヘルプ'. A sidebar on the left contains icons for '目次', '検索', '{x}', and 'ファイル'. The main area is divided into two sections: a 'コードセル' (Code Cell) containing code execution and output, and a 'テキストセル' (Text Cell) containing text and code snippets. The 'コードセル' section is highlighted with a blue border, and the 'テキストセル' section is highlighted with a green border.

Colab の定期購入を最大限に活用する

メニュー

ファイル 編集 表示 挿入 ランタイム ツール ヘルプ

+ コード + テキスト

1. 変数

```
[2]: x = 100
      y = 200
```

2. 式

```
0行: print(x + y)
      print(3 * x + y)
      300
      500
```

3. 条件分岐

```
0行: if (x > 50):
      print('big')
    else:
      print('small')
      big
```

コードセル, テキストセル
の追加

コードセル,
テキストセルの
並び

Web ブラウザの画面

Google Colaboratory のノートブック



コードセル, テキストセルの2種類

- ・**コードセル** : Python プログラム, コマンド, 実行結果
- ・**テキストセル** : 説明文, 図

2. 式

← テキストセル

```
[5] print(x + y)
     print(3 * x + y)
```

← コードセル

300
500

3. 条件分岐

← テキストセル

```
if (x > 50):
    print('big')
else:
    print('small')
```

← コードセル

big

11-1. イントロダクション

人工知能

知的なITシステム

機械学習

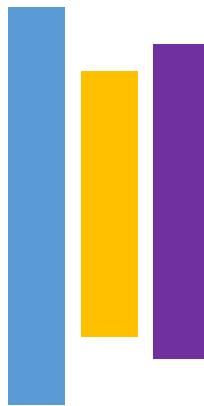
データから**学習**し、知的能力を
向上

ディープラーニング

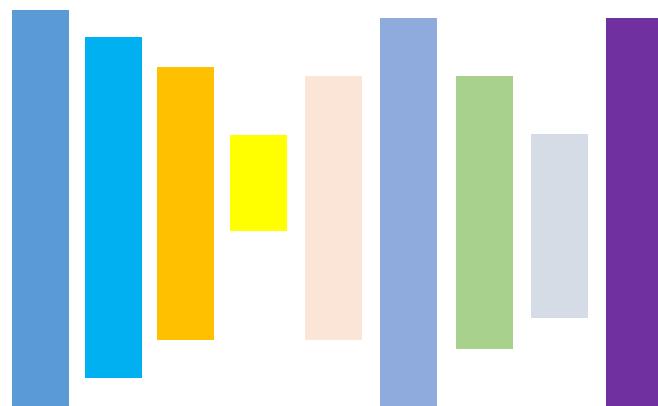
データから**学習**し、複雑なタス
クを実行。**多層のニューラル**
ネットワークを使用

ディープラーニング

ディープラーニングに「ディープ」とついているのは、多層のニューラルネットワークを使用するため



層の数が少ない



層の数が多い（ディープ）

ディープラーニングまとめ



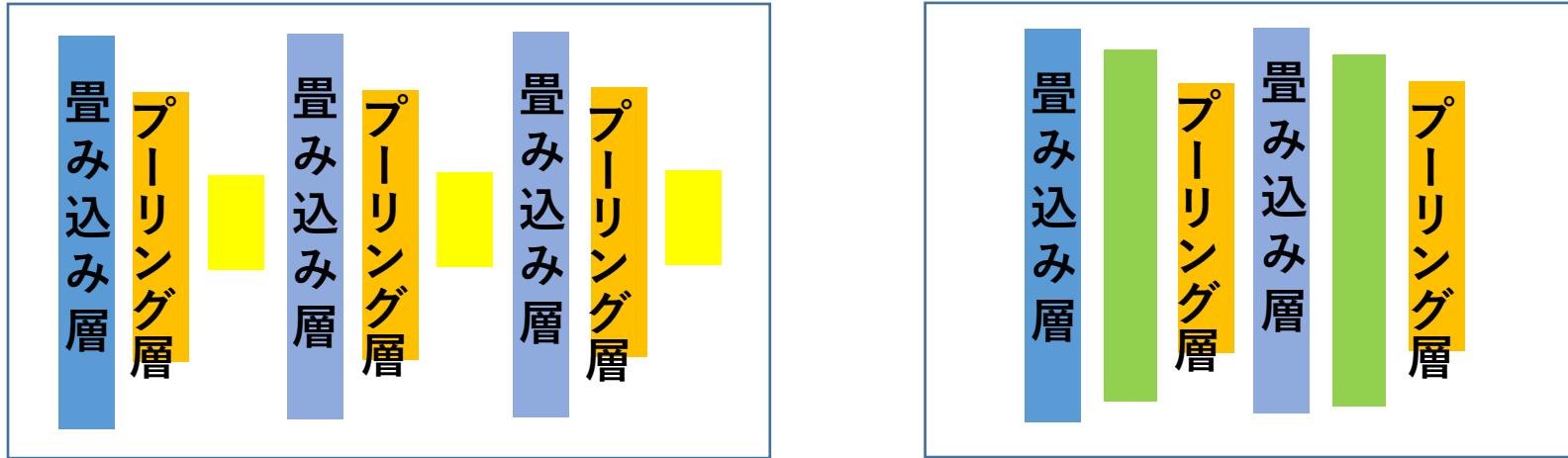
- ・ディープラーニングは機械学習の一種であり、人工ニューラルネットワークを使用してデータから学習し、複雑なタスクを実行する技術
- ・「ディープ」の名前は、多層のニューラルネットワークを使用することに由来
- ・ディープラーニングが広く利用される理由は、多様なデータに適用でき、さまざまなタスクで高性能を発揮するため。例：画像認識、自然言語処理、音声認識など。

畳み込みニューラルネットワーク (CNN)



- ・ 畳み込みニューラルネットワーク (CNN) は画像理解や画像の分析に特化したディープラーニングの一種。
- ・ CNNは主に畠み込み層、プーリング層、全結合層の3種類
(注) これら3種類以外のもさまざまある
- ・ 畠み込み層：画像の局所的な特徴をとらえるための層。特徴は、画像内の顕著なパターンや属性（例：エッジ、テクスチャ）
- ・ プーリング層：特徴マップのサイズを縮小するための層。
過学習を防止。計算効率を向上
- ・ 全結合層：全ニューロンが前の層のすべてのニューロンと接続された層。畠み込み層とプーリング層を通過した後の特徴を基に、画像の分類や回帰を行う

畳み込みニューラルネットワーク (CNN) の例



さまざまなバリエーション

畳み込みニューラルネットワークでのパターン認識

「畳み込みニューラルネットワークの利用により、さまざまなレベルのパターンを抽出・認識できるようになる」という考える場合も

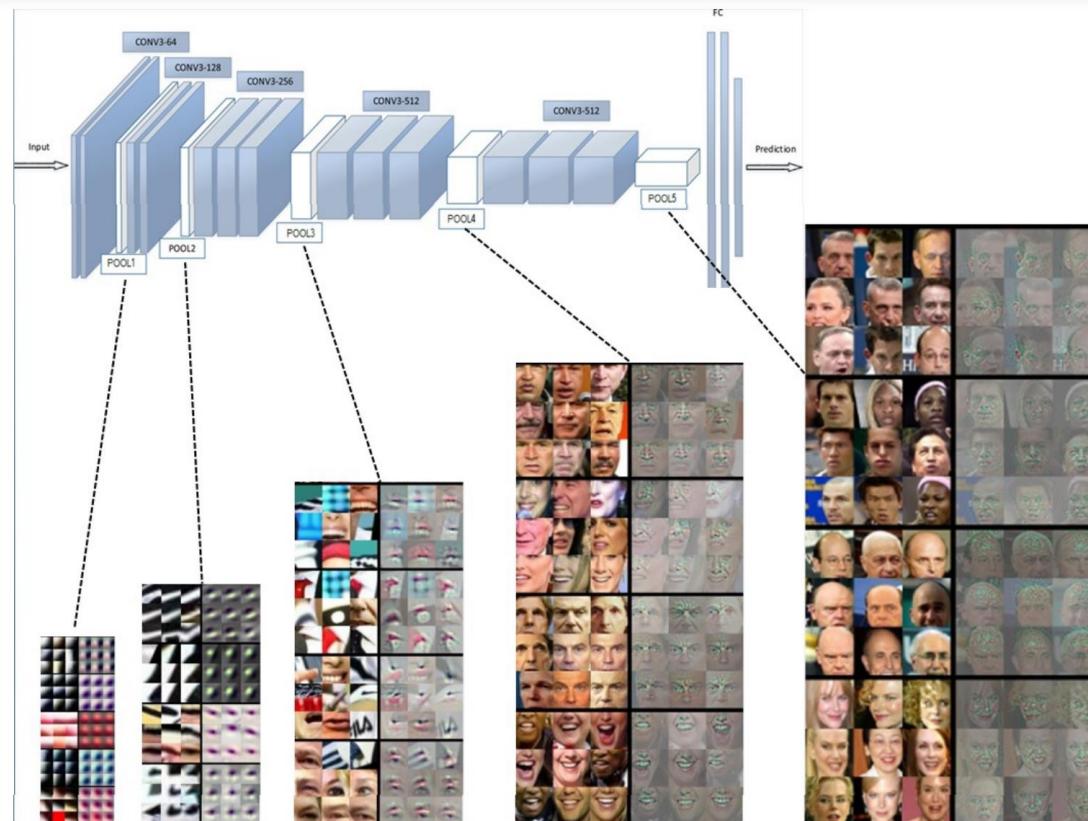


Fig. 2

畳み込みニューラルネットワーク (CNN) のメリット

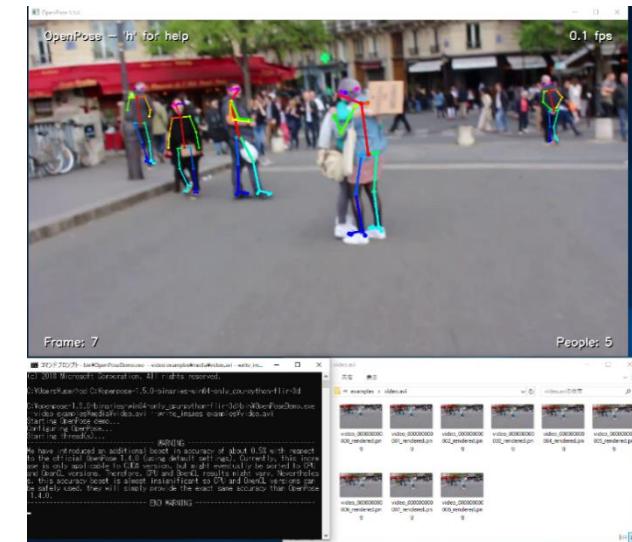


- ・ 画像理解、画像の分析の高精度化
- ・ 画像の特徴を自動的に抽出
- ・ 幅広い分野での応用

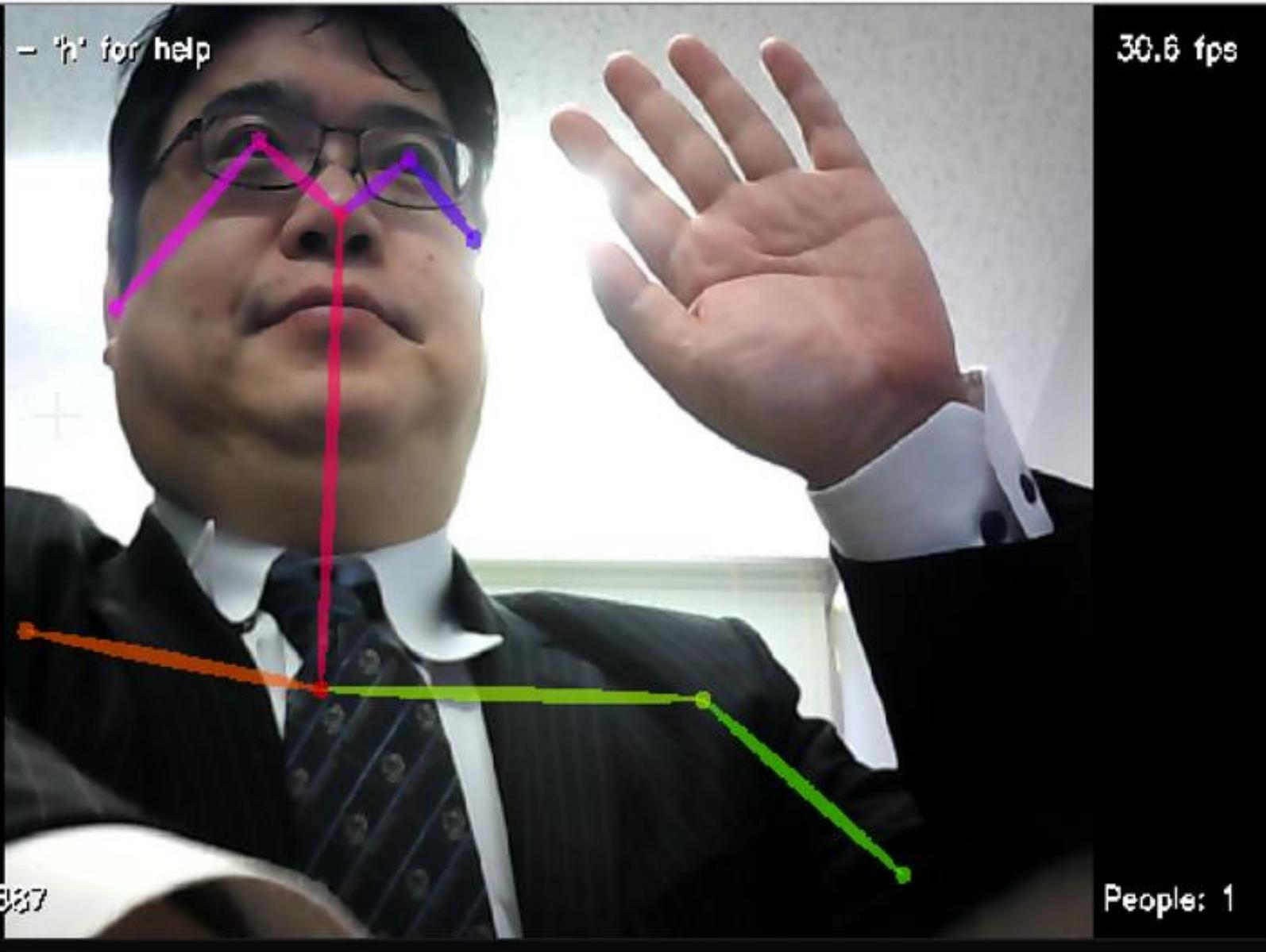
11-2. 姿勢推定の基本

姿勢推定

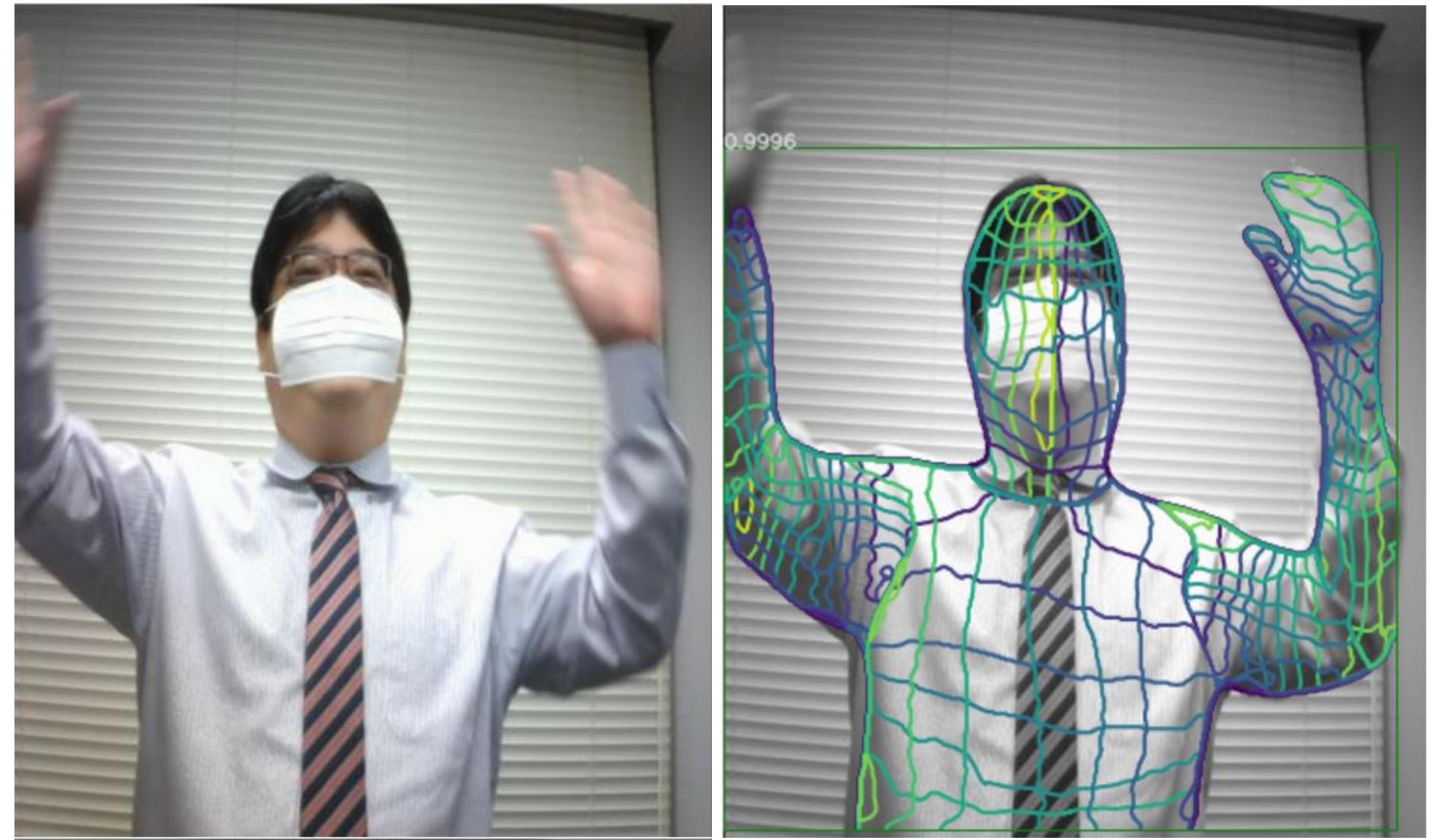
姿勢推定は、人間の全身（人体）、頭部、およびその他のオブジェクトの**位置と方向**を推定する技術



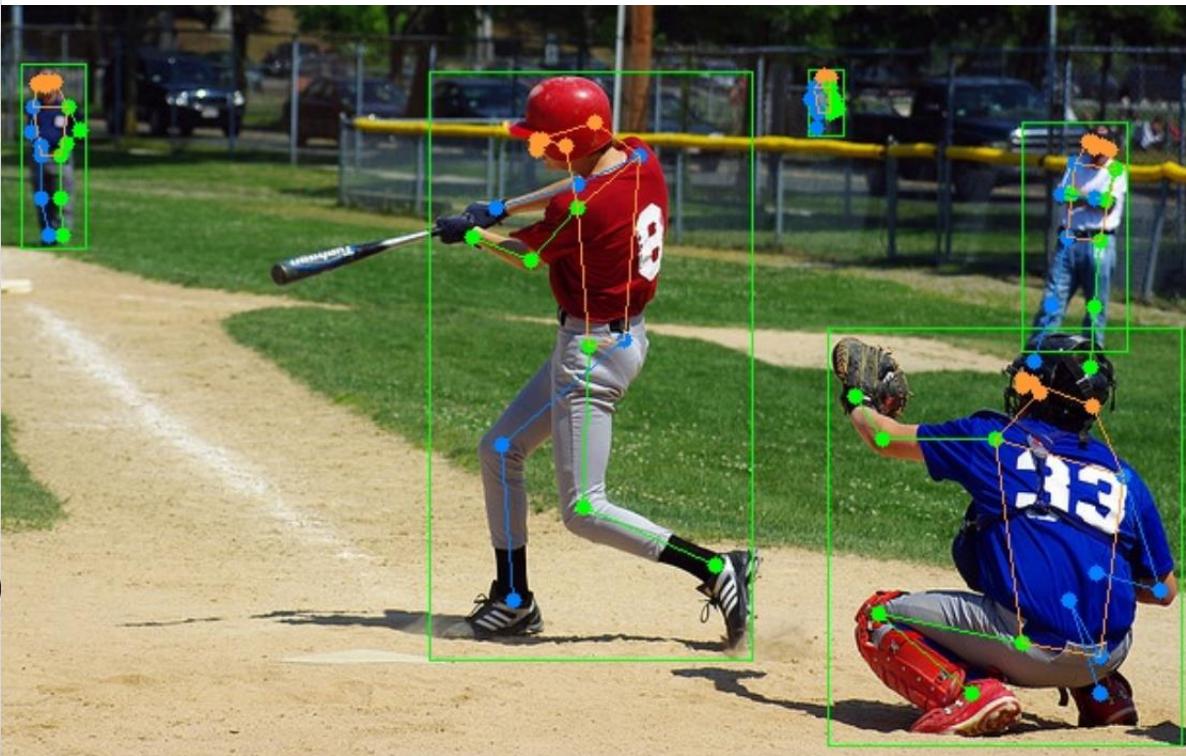
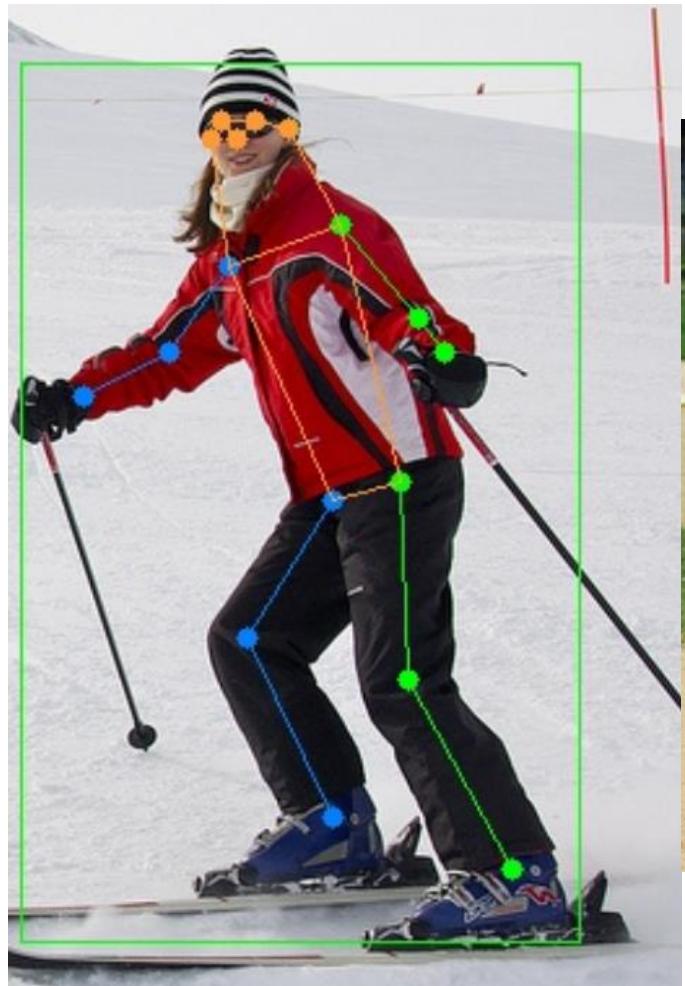
姿勢推定の結果



(OpenPose を使用)



(Detectron2 の DensePose を使用)



ビデオの例



ビデオの例

姿勢推定（人やオブジェクトの動きを読み解く）

- ・ **人間の全身**：人間の動きや行動。その人の活動や意図
- ・ **手や足の指**：手指や足指の動き。細やかな指示や意図
- ・ **頭部**：視線や注視の向き。意図や関心の対象
- ・ **オブジェクト**：オブジェクトのふるまい。オブジェクトと人の相互作用。

姿勢推定の歴史

- ・ **数十年前に研究開始**：姿勢推定は数十年前に開始
- ・ **初期の技術**：センサーやマーカーに基づくシステム
- ・ **画像処理技術の進化**：高度な画像処理技術が登場
- ・ **ディープラーニングの影響**：ディープラーニングにより、姿勢推定の精度と効率が向上

姿勢推定の多様な応用分野

- ・**監視とセキュリティ**：公共の場で、人々の行動を監視。危険や異常を検出
- ・**エンターテイメント**：キャラクター・アニメーションの元データを作成
- ・**ヒューマンコンピュータインタラクション**：人間とコンピュータが相互作用するインタラクション
- ・**スポーツ分析**：選手の動きを分析。怪我を予防
- ・**医療、リハビリ**：身体の動きを評価。リハビリ効果の確認
- ・**頭部追跡の応用**：VR/AR、運転者の集中の確認など
- ・**オブジェクトの向きの応用**：ロボット、倉庫の荷物、自動運転車など

姿勢推定の用途①

- 状況把握，行動予測



姿勢推定の用途②

- 行動検知



姿勢推定の用途③

- 危険の察知，救援・救護



姿勢推定の用途④

- 監視，防犯



姿勢推定の用途⑤

- ・トレーニング，フィットネス，エクササイズ



姿勢推定の用途⑥

- アバター（バーチャル・キャラクタ）の操作



GAN (敵対性生成ネットワーク) のニュース  

- ・**実在しない人間の顔画像を生成**
tl-GAN, https://docs.google.com/presentation/d/1OpcYLBVpUF1L-wwwPHu_CyKjXqXD0oRwBoGP2peSCRSA/edit#slide=id.g4551faa5ed_0_208 より
- ・色分け図や線画をリアルに**変換**
Video-to-Video Synthesis, <https://www.youtube.com/watch?v=S1OwOd-war8> より

アバター（バーチャル・キャラクタ）

姿勢推定の用途⑦

- ・集団行動の観察，分析，監視

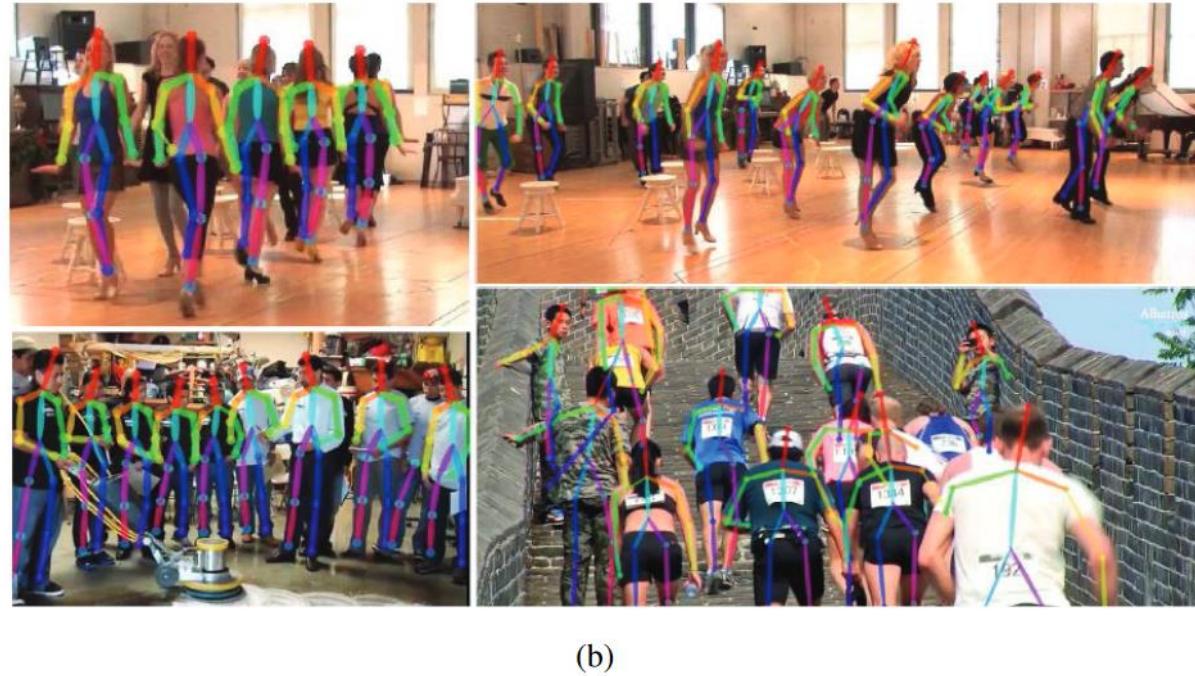


Fig. 1 Examples of pose estimation results. (a) Single person pose estimation results from Ref. [5]. (b) Multi-person pose estimation results from Ref. [3].

Qi Dang, Jianqin Yin, Bin Wang, Wenqing Zheng,

Deep Learning Based 2D Human Pose Estimation: A Survey
Tsinghua Science and Technology
Volume 24 Issue 6 Article 5, 2019.

姿勢推定の用途⑧

- 姿勢推定の結果を用いて、画像から人物を切り出す



ここまでまとめ

姿勢推定は、人間の全身（人体）、頭部、およびその他のオブジェクトの位置と方向を推定する技術

要素

- ・ **人間の全身**：人間の動きや行動。その人の活動や意図。
- ・ **手や足の指**：手指や足指の動き。細やかな指示や意図。
- ・ **頭部**：視線や注視の向き。意図や関心の対象。
- ・ **オブジェクト**：オブジェクトのふるまい。オブジェクトと人の相互作用。

歴史

- ・ 数十年前に始まった。現在は、ディープラーニングの応用が進んでいる

応用分野

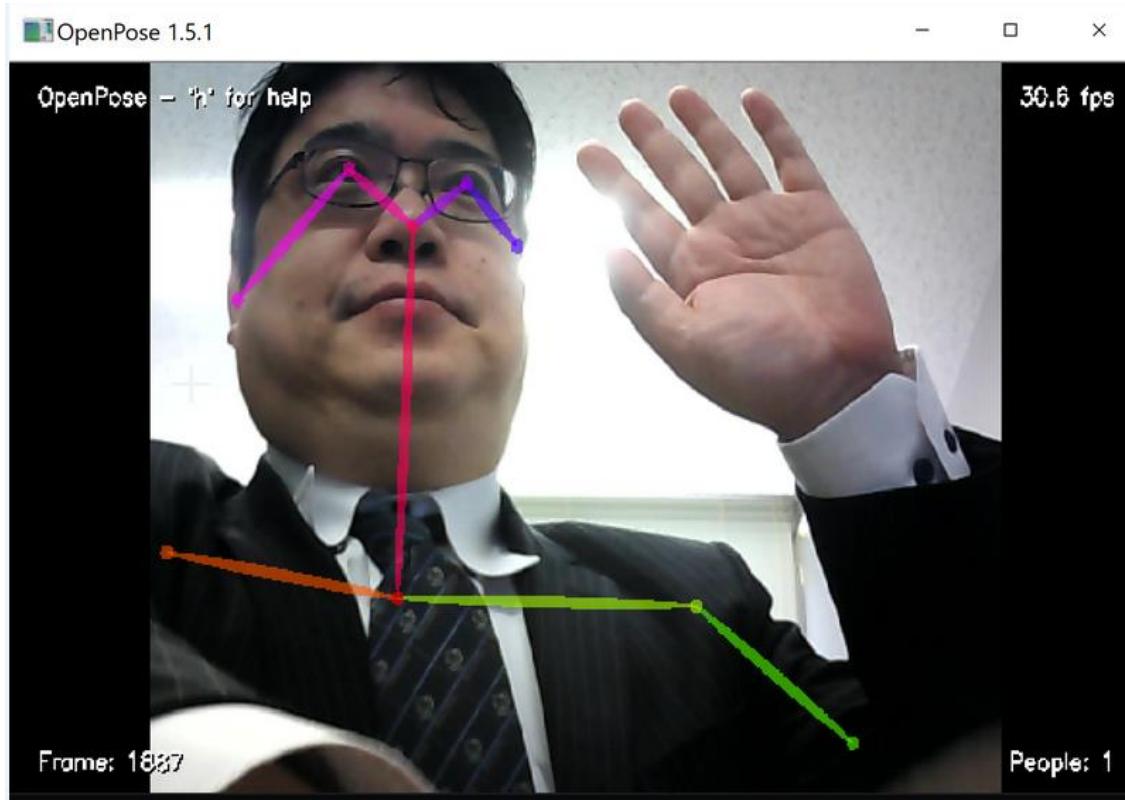
- ・ エンターテイメント、セキュリティ、ヒューマンコンピュータインターラクションなどさまざま

11-3. 人体の姿勢推定

人体の姿勢推定とキーポイント

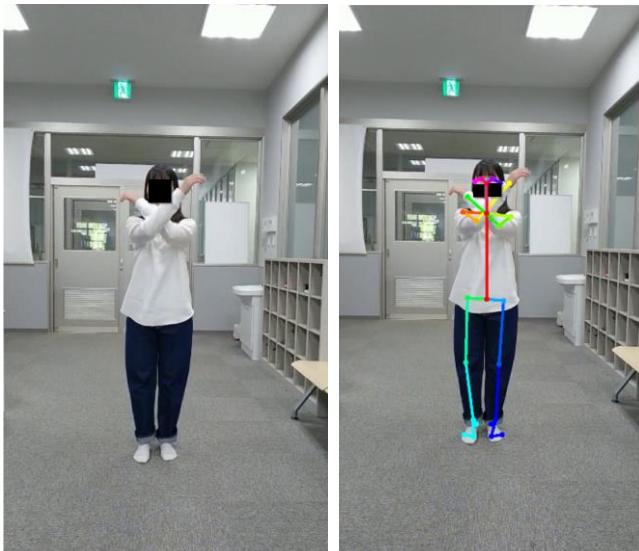
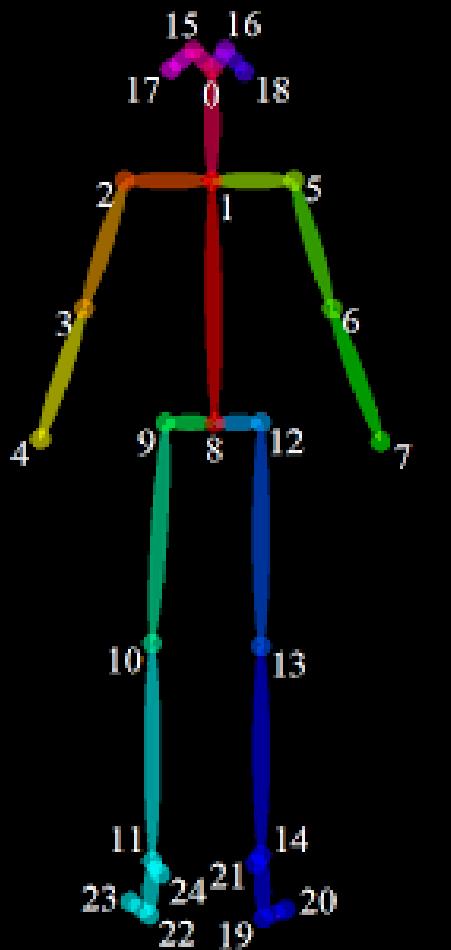


1. キーポイントの検出：姿勢推定では、人体の特定の「キーポイント」（特徴となるポイント）を検出し、位置を特定
2. 姿勢推定：位置が特定されたキーポイントを結びつけることで、人間の全身の姿勢を推定



キーントとボーン

ボーン = 骨格, キーント = 特徴となるポイント



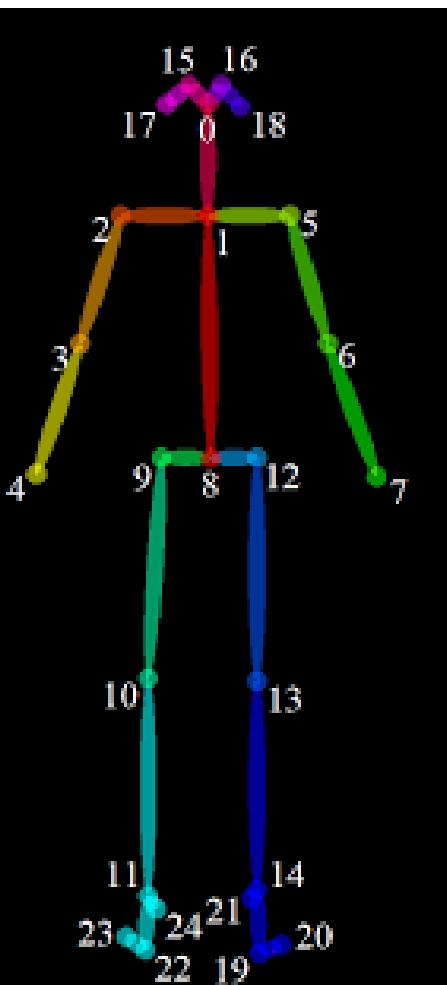
```
"version": 1.3,  
"people": [  
  {  
    "person_id": -1,  
    "pose_keypoints_2d":  
      [363.314, 385.359,  
       432.985, 1008.49,  
       0.979106, 0.888736,  
       363.37, 349.747,  
       506.285, 341.999,  
       0.99886, 0.967798,  
       304.656, 377.803,  
       502.51, 421.022,  
       0.909367, 0.956557,  
       337.621, 333.997,  
       524.568, 429.433,  
       0.840333, 0.802341,  
       418.283, 396.402,  
       493.005, 433.016,  
       0.594983, 0.807274,  
       422.06, 374.436,  
       513.559, 1048.78,  
       0.882184, 0.816693,  
       392.689, 399.931,  
       535.576, 1045.14,  
       0.725604, 0.658681,  
       315.785, 374.45,  
       465.92, 1019.45,  
       0.342554, 0.720292,  
       363.233, 330.279,  
       715.201, 1048.79,  
       0.848111, 0.720292,  
       315.755, 330.279,  
       711.596, 1048.79,  
       0.801404, 0.680071,  
       312.162, 0.860071,  
       689.036, 308.235,  
       0.876103, 1044.99,  
       330.276, 0.892973,  
       1008.47, 337.594,  
       0.892096, 1019.47,  
       403.621, 0.761588,  
       718.853, 0.792035,  
       0.792035, 0.792035,  
       392.782, 0.792035,  
       872.737, 0.792035,  
       0.901708, 0.792035]
```

画像で表示した
ボーン情報

キーントのデータファイル
(特徴点の x, y 座標と
個々の信頼度のデータ)

キーントとボーン

ボーン = 骨格, キーント = 特徴となるポイント



OpenPose では、25個のキーントが定まっている

(OpenPose は、人体の姿勢推定の一手法)

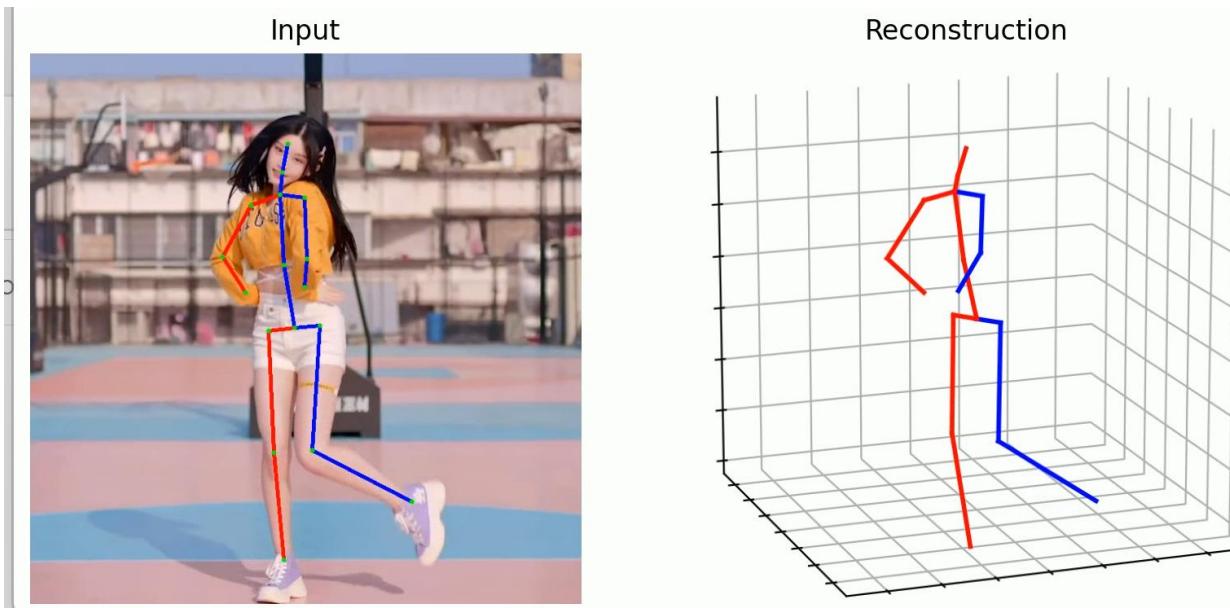
- 0: 鼻 1: 首 2: 右肩 3: 右ひじ 4: 右手首
- 5: 左肩 6: 左ひじ 7: 左手首 8: 真ん中腰
- 9: 右腰 10: 右ひざ 11: 右足首 12: 左腰
- 13: 左ひざ 14: 左足首 15: 右目 16: 左目
- 17: 右耳 18: 左耳 19: 左足親指 20: 左足小指
- 21: 左かかと 22: 右足親指 23: 右足小指
- 24: 右かかと

2次元と3次元の姿勢推定

姿勢推定は2次元（平面的な推定）と3次元（空間的な推定）の2種類

- 2次元: 2次元的な（平面の）姿勢を推定
- 3次元: 空間的な深さ情報（奥行き）を含め、3次元的な姿勢を推定

3次元の姿勢推定は、より詳細な情報を提供



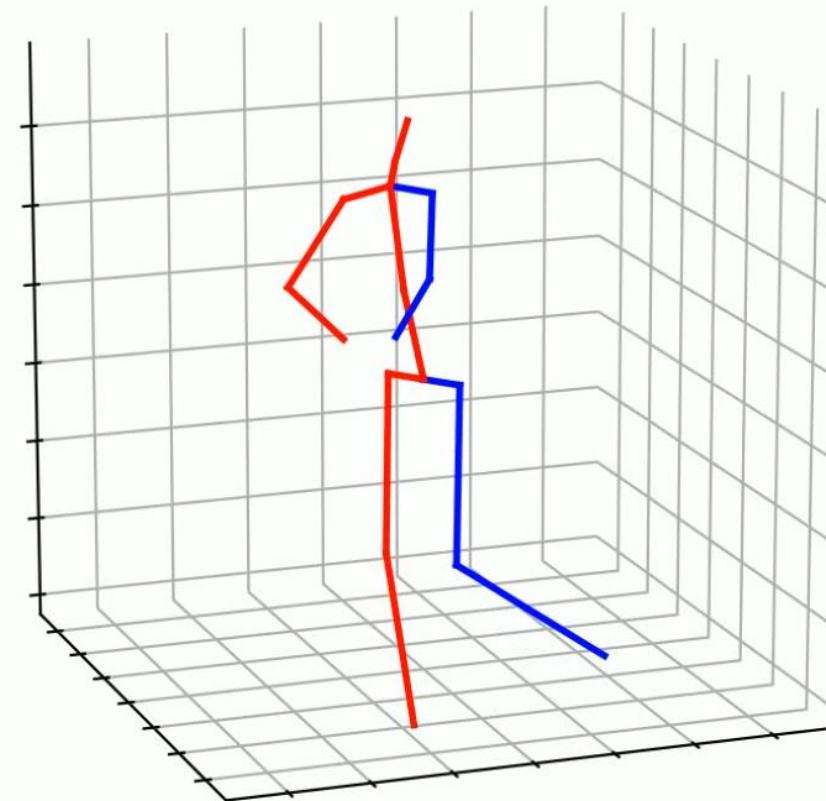
3次元姿勢推定

3次元の姿勢推定

Input



Reconstruction

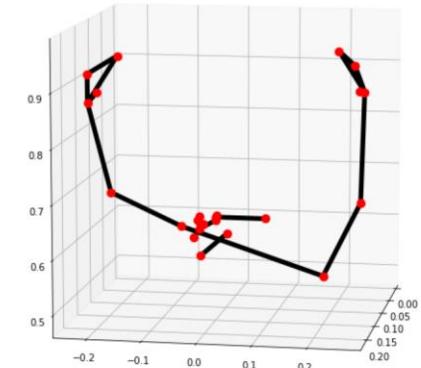


- 3次元の姿勢推定は、空間的な深さ情報（奥行き）を含め、3次元的な姿勢を推定
- 人体表面のポリゴンメッシュを付けたり、アクションの分類を行うなどの応用

3次元の姿勢推定の例



キーポイントとボーン



キーポイントの位置情報
は3次元（縦、横、奥行き）

ここまでまとめ

人体の姿勢推定とキーポイント

- ・ **姿勢推定**では、人体の特定の「キーポイント」（特徴となるポイント）を検出し、**位置を特定**
- ・ **姿勢推定**：位置が特定されたキーポイントを結びつけることで、人間の全身の姿勢を推定
- ・ OpenPose：人体の姿勢推定の一手法。25個のキーポイント（鼻、首、右肩、右ひじなど）

2次元と3次元の姿勢推定

- ・ **2次元の姿勢推定**：2次元的な（平面の）姿勢を推定
- ・ **3次元の姿勢推定**：空間的な深さ情報（奥行き）を含む、より詳細な姿勢推定

11-4. ディープラーニング による姿勢推定の仕組み

姿勢推定とディープラーニング



ディープラーニングの利用により、姿勢推定の精度が向上。

以下のような課題が部分的に解決

- ・多人数の対応
- ・人体の向きや姿勢のバリエーション
- ・個々の人体の個性
- ・部分遮蔽
- ・照明のバリエーション（夜間、影）
- ・衣服のバリエーション

さまざまな状況において、部分隠蔽があっても、姿勢推定ができるようになってきた

姿勢推定とディープラーニング

ディープラーニングは万能ではなく、次のような問題がある

- **訓練データのバリエーションの不足**

姿勢推定のために「制限のある実験環境」で撮影されたデータを使う場合、**訓練データのバリエーションが限られ、その範囲でしか姿勢推定できない**

この問題の解決のため、自己教師あり学習などのアプローチが行われている

ディープラーニングによる姿勢推定

1. キーポイントの検出

次の手順で、複数人体のキーポイント（特徴となるポイント）を識別

- ① **部位の位置推定**: 鼻、首、右肩、右ひじなどの**部位**を検出。
- ② **同一人物のキーポイント特定**: ステップ①の結果とともに、**同一人物の部位**を集めて、最終的なキーポイントを得る

2. 姿勢推定

位置が特定された**キーポイントを結びつける**ことで、人間の全身の姿勢を推定

キーポイントの検出

元画像



(a)

①部位の
位置推定



(b)

②同一人物の
キーポイント特定



(c)

Fig. 6 An example of heatmap-based single-person pipeline with heatmap. (a) Original image, (b) heatmap generated by estimator, and (c) detection result.

Qi Dang, Jianqin Yin, Bin Wang, Wenqing Zheng,

Deep Learning Based 2D Human Pose Estimation: A Survey
Tsinghua Science and Technology
Volume 24 Issue 6 Article 5, 2019.

部位の位置推定を行う Hourglass

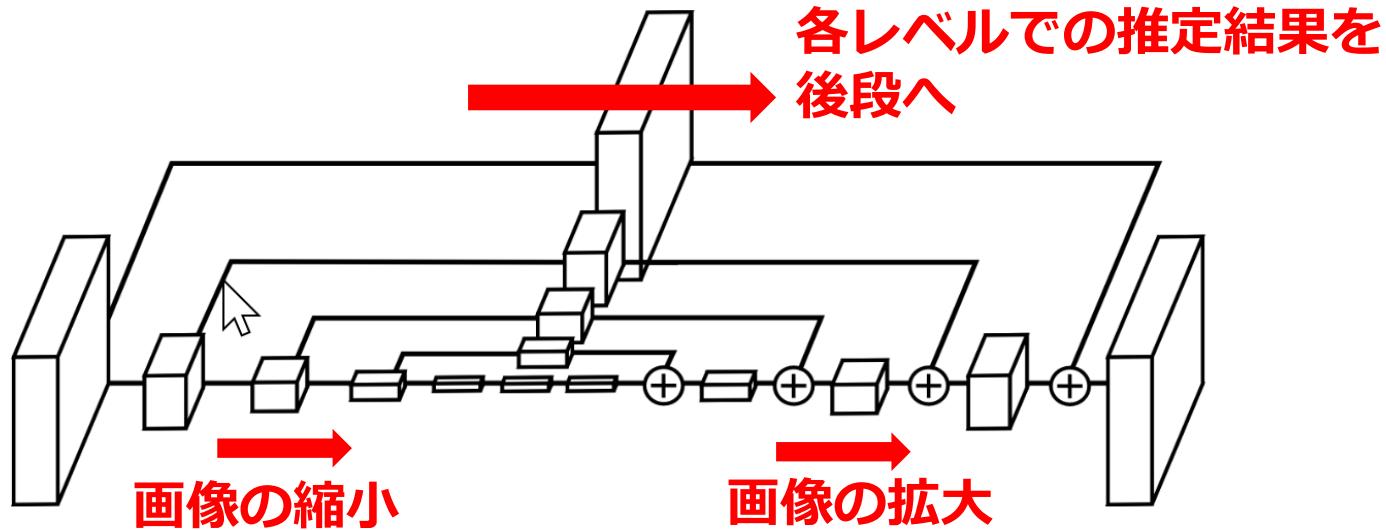


Fig. 3. An illustration of a single “hourglass” module. Each box in the figure corresponds to a residual module as seen in Figure 4. The number of features is consistent across the whole hourglass.

部位の位置推定

部位の大きさはさまざま ⇒ マルチスケールの画像を使う

- 画像を複数レベルに縮小し、各レベルで部位を検出
- 元の大きさに戻しながら、検出結果を重ね合わせ

Alejandro Newell, Kaiyu Yang, and Jia Deng,
Stacked Hourglass Networks for Human Pose Estimation,
arXiv:1603.06937v2, 2016.
<https://arxiv.org/pdf/1603.06937v2.pdf>

部位の位置推定を行う Hourglass

- 画像の縮小： 置み込み層とプーリング層で可能
- 画像の拡大： 逆置み込み層あるいはアップサンプリング層で可能

Hourglass は、部位の位置推定に特化した置み込みニューラルネットワーク

3次元の人体姿勢推定の技術

- MotionAGFormer (2023年)
- MHFormer (2022年)
- MotionBERT (2022年)
- DiffPose (2022年)

ディープラーニングによる3次元の姿勢推定

1. キーポイントの検出

次の手順で、複数人体のキーポイント（特徴となるポイント）を識別

- ① **部位の位置推定**: 鼻、首、右肩、右ひじなどの**部位**を検出。
- ② **同一人物のキーポイント特定**: ステップ①の結果とともに、**同一人物の部位**を集めて、最終的なキーポイントを得る
- ③ ②の結果の**キーポイント**に対して、**空間的な深さ（奥行）**を求め、**3次元の姿勢推定**を実行 = 追加処理

2. 姿勢推定

位置が特定された**キーポイント**を結びつけることで、人間の全身の姿勢を推定

ここまでまとめ

人体の姿勢推定の精度向上

多人数対応、姿勢のバリエーション、部分遮蔽、照明・衣服のバリエーションへの対応。

処理手順

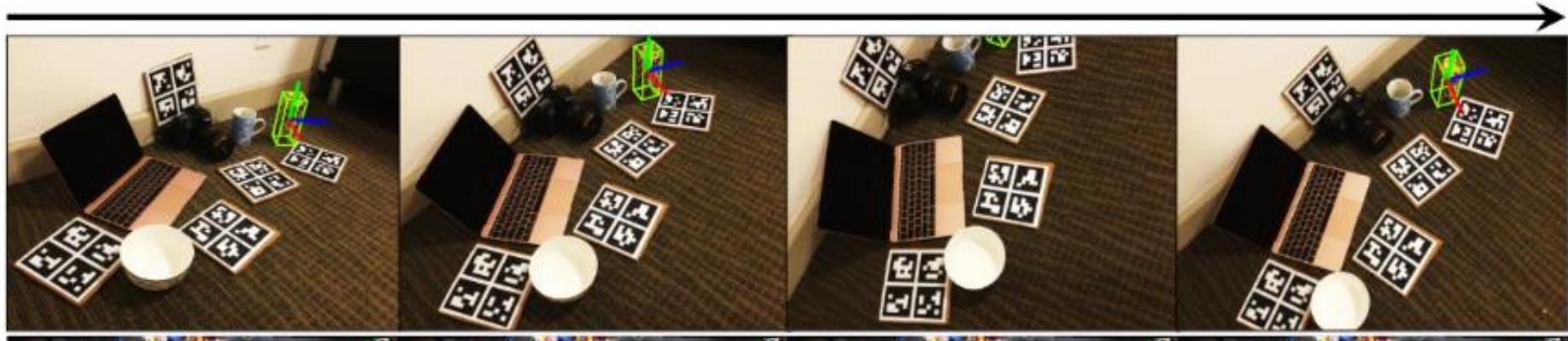
- 部位の位置推定
- 同一人物のキーポイント特定
- キーポイントに対して深さ（奥行）を求める（3次元の姿勢推定を行う場合の追加処理）。

ディープラーニングによる姿勢推定の現在の課題

訓練データのバリエーション不足、自己教師あり学習のアプローチ。

11-5. 頭部の姿勢推定、物 体の姿勢推定

物体の姿勢推定

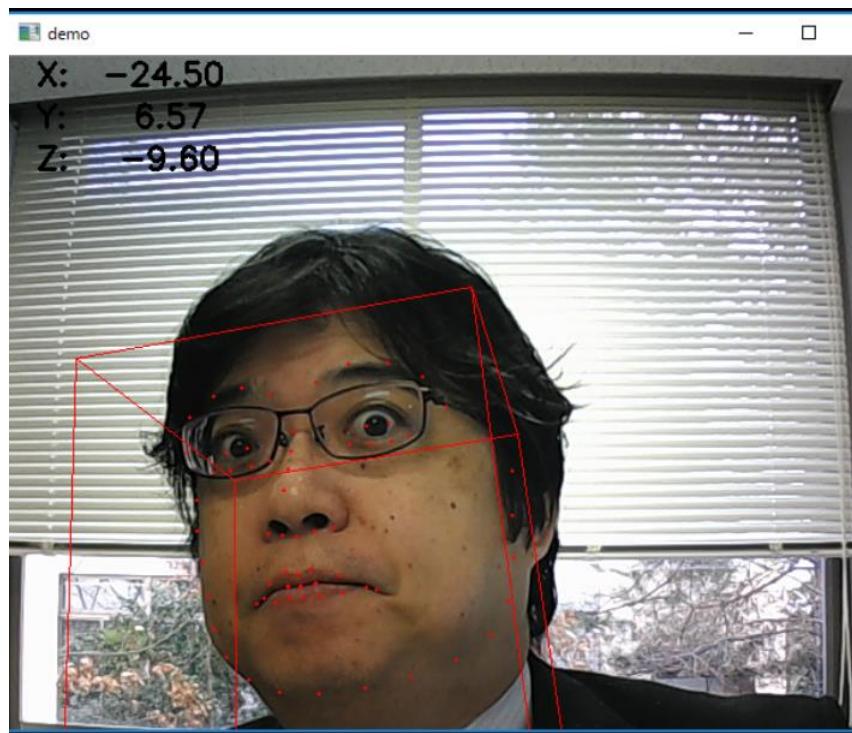


- 画像から、物体の3次元の位置（たて、よこ、奥行き）と方向の情報を得る

Wen, B and Bekris, Kostas E,

BundleTrack: 6D Pose Tracking for Novel Objects without Instance or Category-Level 3D Models,
IEEE/RSJ International Conference on Intelligent Robots and Systems, 2021.

頭部の姿勢推定

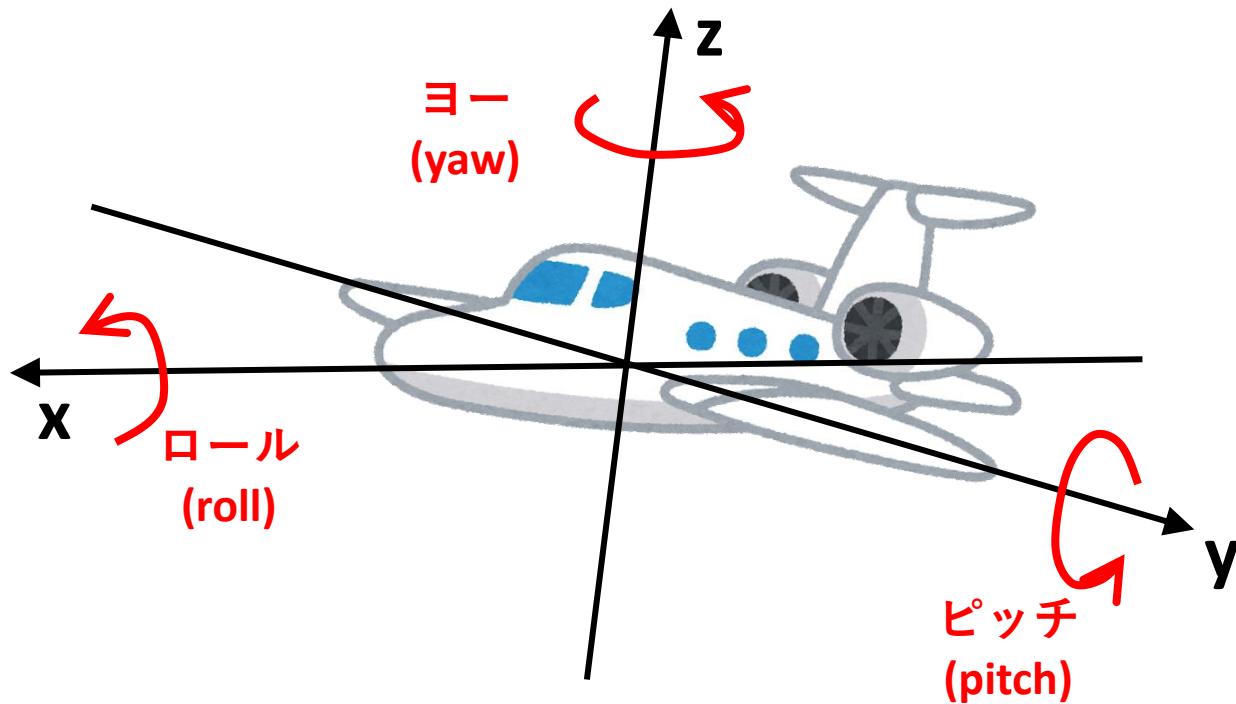


- 頭部の画像から、頭部の**3次元の位置（たて、よこ、奥行き）と方向の情報**を得る

lincolnhard/head-pose-estimation を使用

<https://github.com/lincolnhard/head-pose-estimation>

方向：基準となる3つの軸から、どれだけ回転したか



物体の位置推定，頭部の位置推定

- 画像から，物体や頭部の次の結果を得る
位置（縦，横，奥行きの**3つの数値**）
向き（3つの回転軸なので，**3つの数値**）

物体の姿勢推定の用途

- ・ロボットハンドによる物体移動

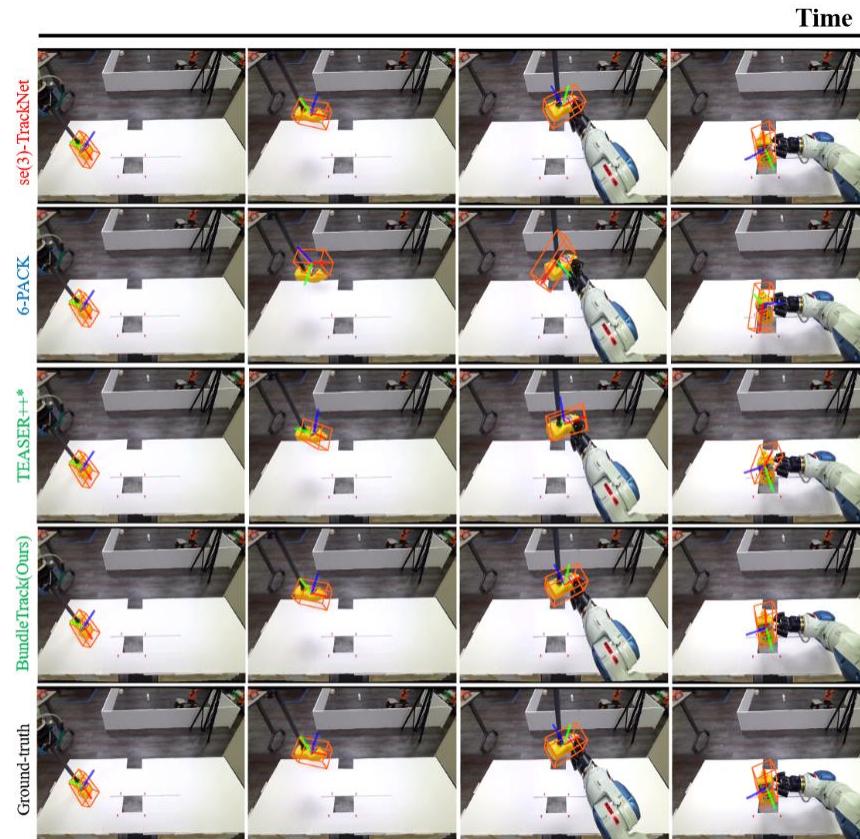


Fig. 4: Example qualitative results of *BundleTrack* and representative comparison points on *YCBInEOAT Dataset*. Methods' names are colored in red, blue and green to denote assumption on *instance 3D model*, *category-level 3D model* and *no model* respectively. For more qualitative results, please refer to the supplementary video.

Wen, B and Bekris, Kostas E,

BundleTrack: 6D Pose Tracking for Novel Objects without Instance or Category-Level 3D Models,
IEEE/RSJ International Conference on Intelligent Robots and Systems, 2021.

物体の姿勢推定

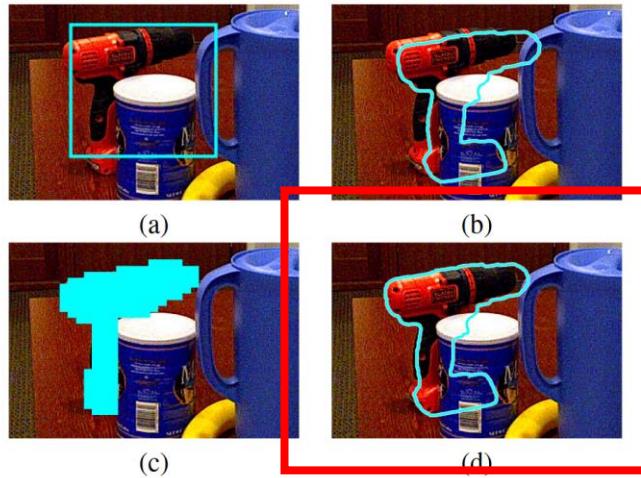


Figure 1: **Global pose estimation vs our segmentation-driven approach.** (a) The drill's bounding box overlaps another occluding it. (b) As a result, the globally-estimated pose [45] is wrong. (c) In our approach, only image patches labeled as corresponding to the drill contribute to the pose estimate. (d) It is now correct.

- 部分隠蔽された物体についても、姿勢推定（位置と方向の推定）ができるようになる
(姿勢推定により、セグメンテーションの精度が向上する、という考え方も登場)

11-6. 演習

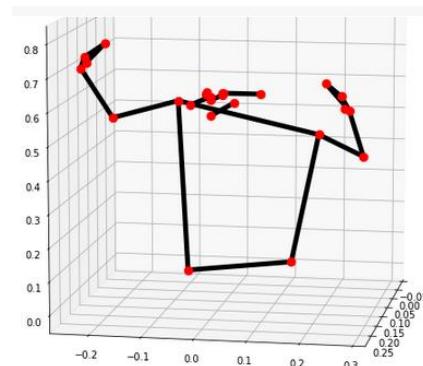
人体の3次元姿勢推定①

1. 使用するページ:

https://colab.research.google.com/drive/13nOMSWODzx_LjN9XEG99jtvgMACI4m9V

2. 目的

姿勢推定、キーポイント、ボーンについて理解を深める



3. 各自分で行うこと

各自で説明、ソースコード、実行結果を確認する。

人体の3次元姿勢推定②

1. 使用するページ:

https://colab.research.google.com/drive/1QRLqEAePmgS41vOKQwf87G_Ss_BLhPYs?usp=sharing#scrollTo=nSbrUG6xxz7r

(LART の研究者によるデモページ)

2. 目的

LART の作者によるデモページにより、

3次元の姿勢推定の理解を深める

3. 各自分で行うこと

各自で説明、ソースコード、実行結果を確認する。



クリックして
展開

> Visualize the reconstructions

▶ ↓ 1 個のセルが非表示

全体まとめ①

姿勢推定は、人間の全身（人体）、頭部、およびその他のオブジェクトの位置と方向を推定する技術

人体の姿勢推定とキーポイント

- ・ **姿勢推定**では、人体の特定の「キーポイント」（特徴となるポイント）を検出し、**位置を特定**
- ・ **姿勢推定**：位置が特定されたキーポイントを結びつけることで、人間の全身の姿勢を推定

2次元と3次元の姿勢推定

- ・ **2次元の姿勢推定**：2次元的な（平面の）姿勢を推定
- ・ **3次元の姿勢推定**：空間的な深さ情報（奥行き）を含む、より詳細な姿勢推定

全体まとめ②

ディープラーニングによる人体の姿勢推定

人体の姿勢推定の精度向上

多人数対応、姿勢のバリエーション、部分遮蔽、照明・衣服のバリエーションへの対応。

処理手順

- 部位の位置推定
- 同一人物のキーポイント特定
- キーポイントに対して深さ（奥行き）を求める（3次元の姿勢推定を行う場合の追加処理）

物体と頭部の姿勢推定

位置と方向の情報を画像から得る