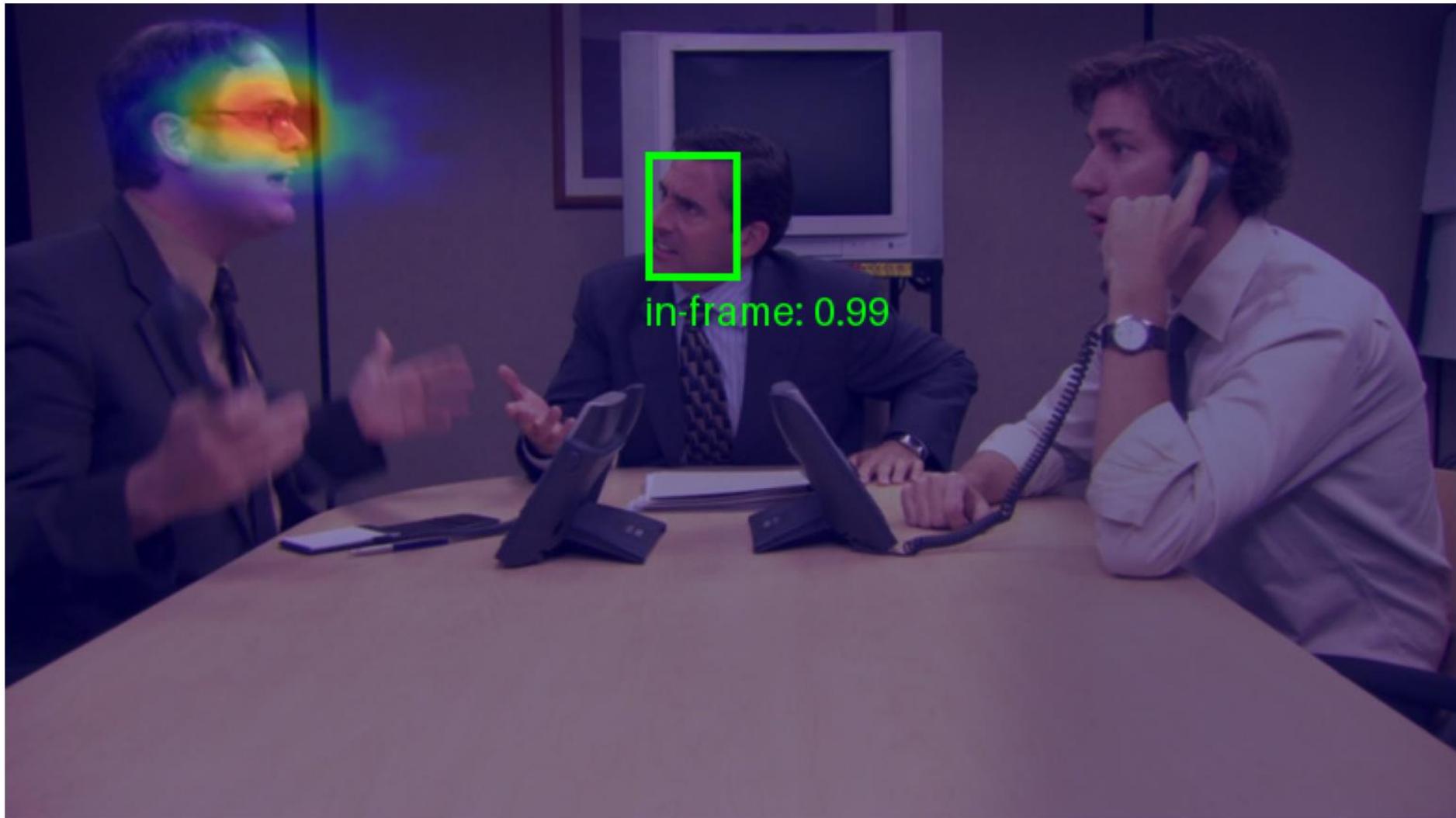


# Gaze-LLE による視線 推定

# Person 1 Gaze Heatmap



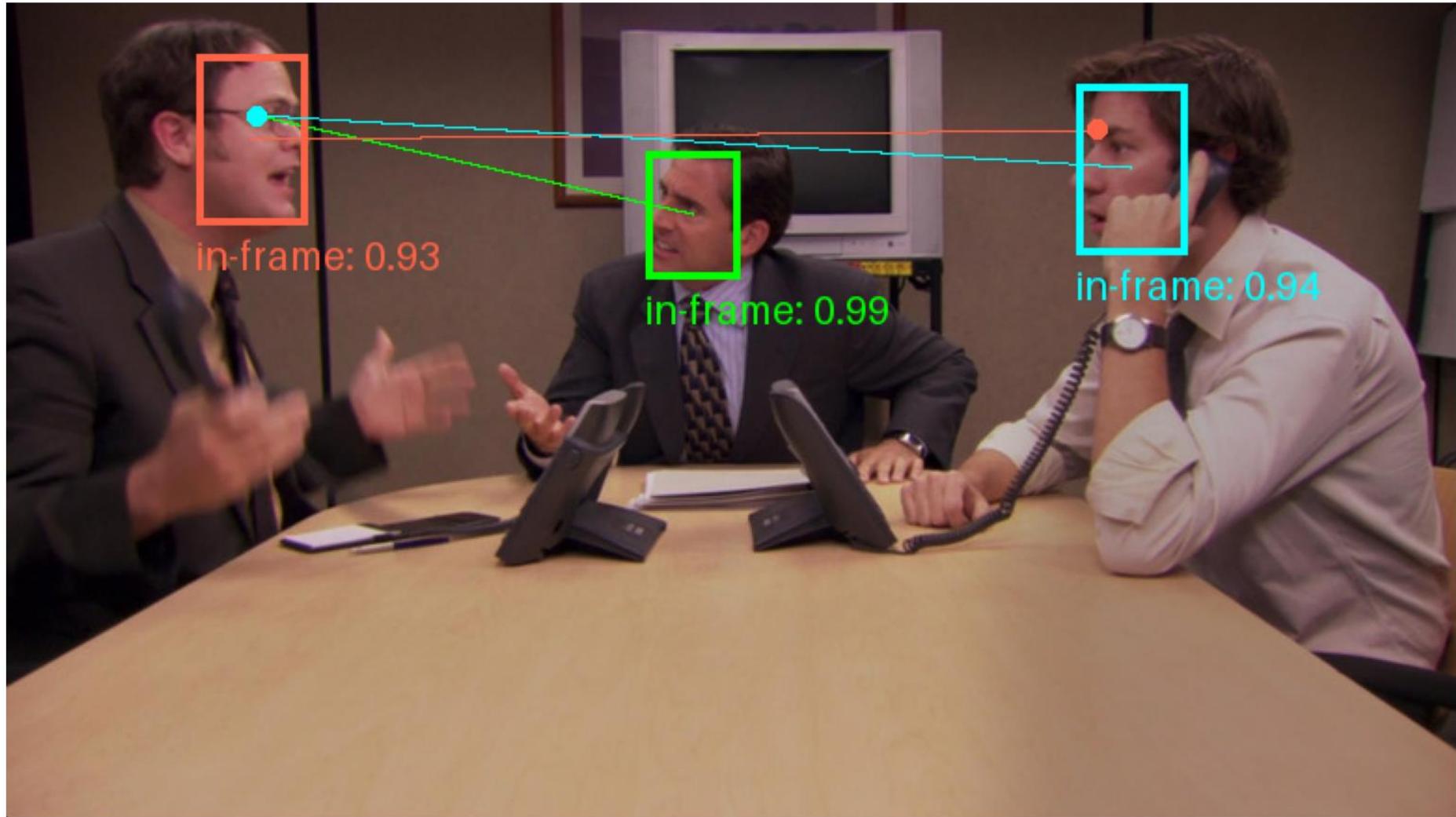
## Person 2 Gaze Heatmap



### Person 3 Gaze Heatmap



# Combined Gaze Visualization



# 視線推定とは

- ・画像内的人物がどの位置を見ているかを特定

## 技術的課題

- ・視線方向は**頭部の向きと目の動き**から決まる。
- ・加えて、**シーン全体のどこに注目対象があるか**を推定する。

## Gaze-LLEの解決方法

- ・技術の組み合わせ

1. **顔検出**→専用のAI (RetinaFace)
2. **画像全体の理解**→DINOv2が担当 (シーン構造、物体配置を認識)
3. **視線方向の推定**→DINOv2 とは別の学習済身  
ニューラルネットワーク

# DINOv2の役割

## 画像認識における特徴抽出

- 画像はピクセルの集まり
- 特徴抽出は、ピクセル値から「人物の位置」「物体の配置」「空間の構造」といった意味情報を取り出す処理である。

## DINOv2の性能

- 1億枚以上の画像で学習済み
- 物体やシーン構造を認識

## Gaze-LLEでの使用方法

- DINOv2が output する 特徴を 基に 視線を 計算する。

# RetinaFaceによる顔検出

## 顔検出が必要な理由

- ・画像内に複数人がいる場合、各人物を区別する必要がある。
- ・また、**視線は頭部と目の位置関係から決まる。**頭部位置は必要な情報である。

## RetinaFaceの出力

- ・画像内の全顔領域
- ・各顔の矩形座標
- ・目、鼻、口の5点座標

## 検出結果の利用方法

- ・顔座標をGaze-LLEに与える

# 全体の処理フロー

## ステップ1：対象人物の特定

- RetinaFaceが画像から顔を検出し、各人物を識別する。

## ステップ2：シーン理解

- DINOv2が画像全体から特徴を抽出し、物体配置や空間構造の情報を取得する。

## ステップ3：視線計算

- Gaze-LLEが顔位置と抽出された特徴を統合し、各人物の視線方向を計算する。

## ステップ4：結果出力

- $64 \times 64$ の確率分布として出力する。各マスの値は、その位置が注視点である確率を表す。