

# 個別ゼミ

2025 0909 工藤混青

# アウトライン

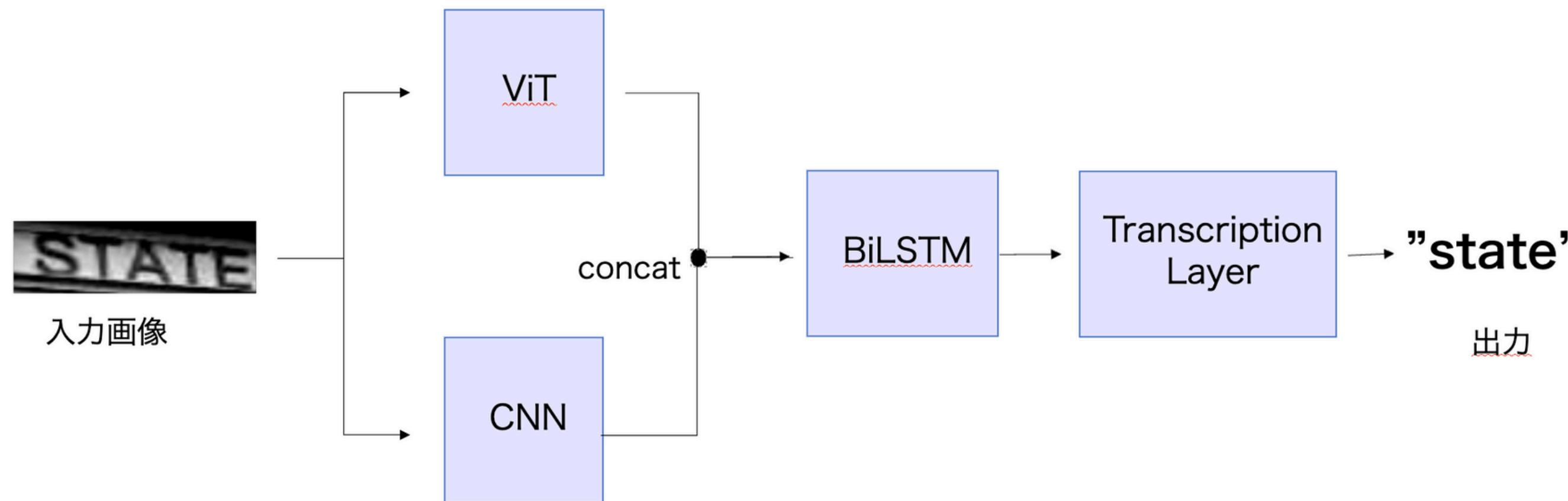
1. 中間報告の振り返り
2. やってきたこと
3. 今後の展望

# 中間発表の振り返り

- 質問につまつた集
  - 新規性はどこ？？ by 明先生
- 発表の説明は社会的意義って大事？？
  - 基礎研究でそれってどうやって出すの？
  - 歴史資料の保存・検索、行政や業務の効率化、教育・アクセシビリティ向上を促す社会的基盤技術的な？
  - また、自分が新たにHTRをするのではなくすでにあるものを精度向上をさせるものの場合上記の感じでいいのか
- 研究とは

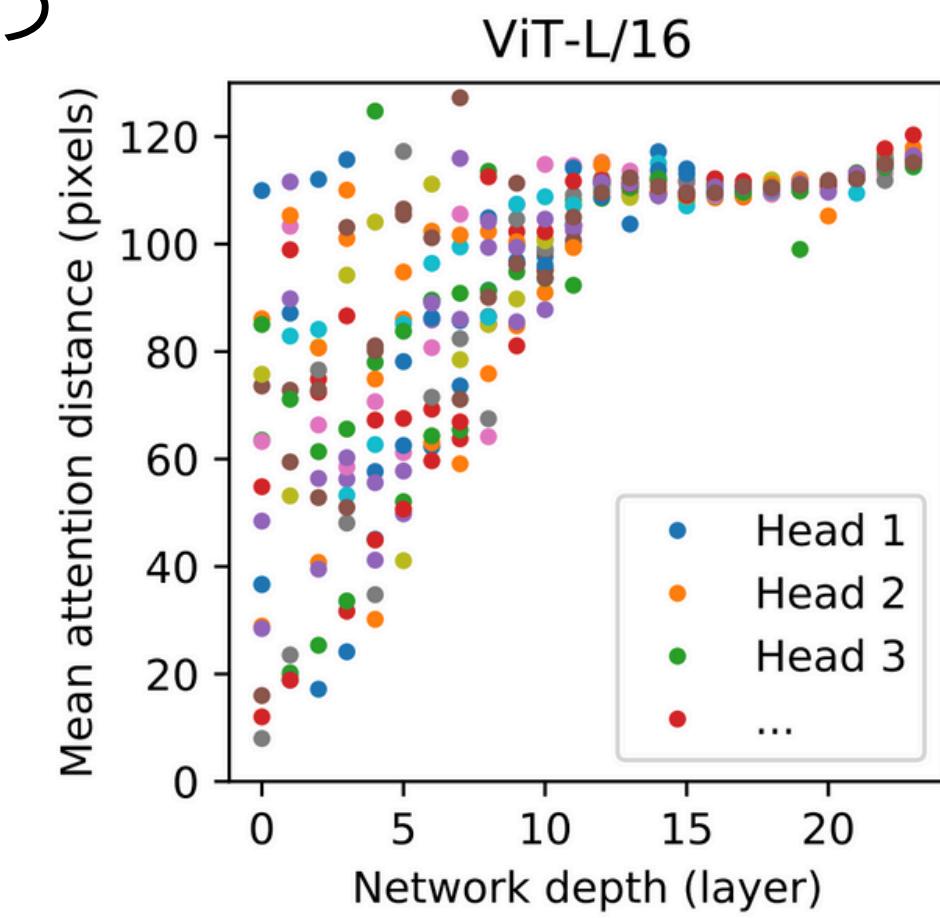
# やってきたこと

- <https://github.com/meijieru/crnn.pytorch>の編集
- ViTを組み込んだコードの理解
  - 次ページに詳細



# これまでやってきたこと

- 現状のコードの整理
  - 位置埋め込み
    - sinusoidal position embedding
  - ViTを途中(epoch数が約20%の段階)で解凍
    - 学習初期段階ではグローバル特徴取れなさそう
    - ここでも比較実験を一応やった方がいい？
  - 最適化 : Adam W
  - Loss : CTCLoss
  - epoch\_num : 40
  - batch\_size : 16
  - 訓練データ : 9912
  - 検証データ : 1102



ViT論文より

# これまでやってきたこと

- 現状のモデルの推論力
  - 良い面
    - 単語や 1 文単位かつ学習用データと同じデータ合成方法なら大文字小文字を考慮しない場合認識精度が非常に高い
  - 課題
    - 学習データと同じ画像合成方法でない画像への推論力低下
    - 大文字小文字区別

Turner, who made his money in Australia and returned some years ago to  
入力画像

turner, who made his money in australia and returned some years ago to  
出力結果

element  
入力画像

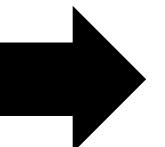
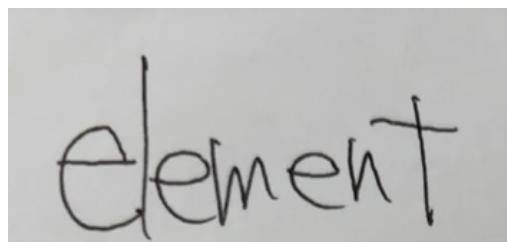
\_v4v  
推論結果

# これまでやってきたこと

- 課題の原因考察
  - a. 推論時に画像サイズが指定のものでない場合強制リサイズ処理が施されそれに  
よって人間ですら読み取りが難しい手書き文字画像になっているため
  - b. データセットが同じ合成方法だから？

# これまでやってきたこと

- 仮説aの検証
  - 先ほどの画像が推論時どのような画像になるのか調査
- 解決策アスペクト比を保ったまま画像を推論してみる



before

output : \_v4v

after

```
if keep_ratio:  
    # アスペクト比保持モード: 画像の歪みを防ぐ  
    # 高さを基準にリサイズし、幅はパディングで調整  
    rsz = T.Resize(imgH)(img)  
    # 黒い背景でパディング用キャンバスを作成  
    pad = Image.new('L', (imgW, imgH), color=0)  
    # リサイズした画像を左上に配置  
    pad.paste(rsz, (0,0))  
    # テンソル化と正規化  
    x = T.ToTensor()(pad)  
    x = T.Normalize((0.5,), (0.5,))(x) # [0,1] -> [-1,1]
```

アスペクト比の保持コード

# これまでやってきたこと

- 仮説bの検証
  - IAMの手書き画像で学習させてみる

# 今後の展望

- データセットのグレードアップ
  - IAMのデータセットをつかてみる
- 他に使えるモデルや手法がないかの調査
-