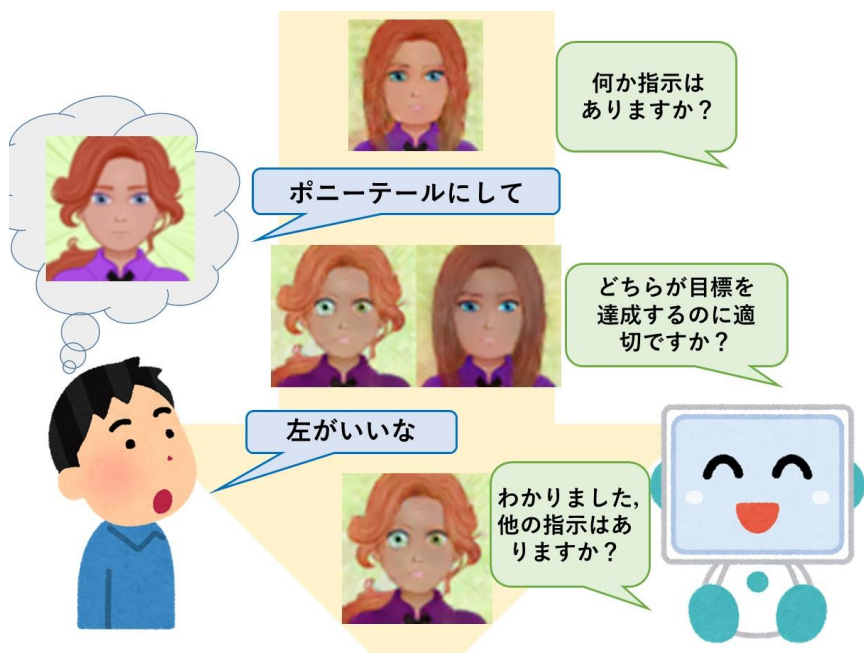


自然言語による編集要求に対して効率的に確認を行う 対話的画像編集システム

本研究の詳しい内容はこちら：<https://ieeexplore.ieee.org/document/9099288>



対話デモ動いています (8/5限定)

URLから：https://t.me/avatar_edit_dial_bot

QRコードから：



Telegram-botと
対話してみよう!

注：Telegram Webで使う場合も、**最初に
スマホでTelegramアカウント作成が必要**です

iOS: Telegram Messenger
Android: Telegram

品川政太郎, 吉野幸一郎, Seyed Hossein Alavi,
Kallirroï Georgila, David Traum, Sakriani Sakti, 中村哲

奈良先端大, 理研AIP, USC ICT

自然言語による編集要求に対して効率的に確認を行う 対話的画像編集システム

とは？

本研究の詳しい内容はこちら：<https://ieeexplore.ieee.org/abstract/document/825258>



対話デモ動いています(8/5限定)
URLから：https://t.me/avatar_edit_dial_bot
QRコードから： Telegram-botと対話してみよう！

注：Telegram Webで使う場合も、**最初にスマホでTelegramアカウント作成が必要**です

iOS: Telegram Messenger
Android: Telegram

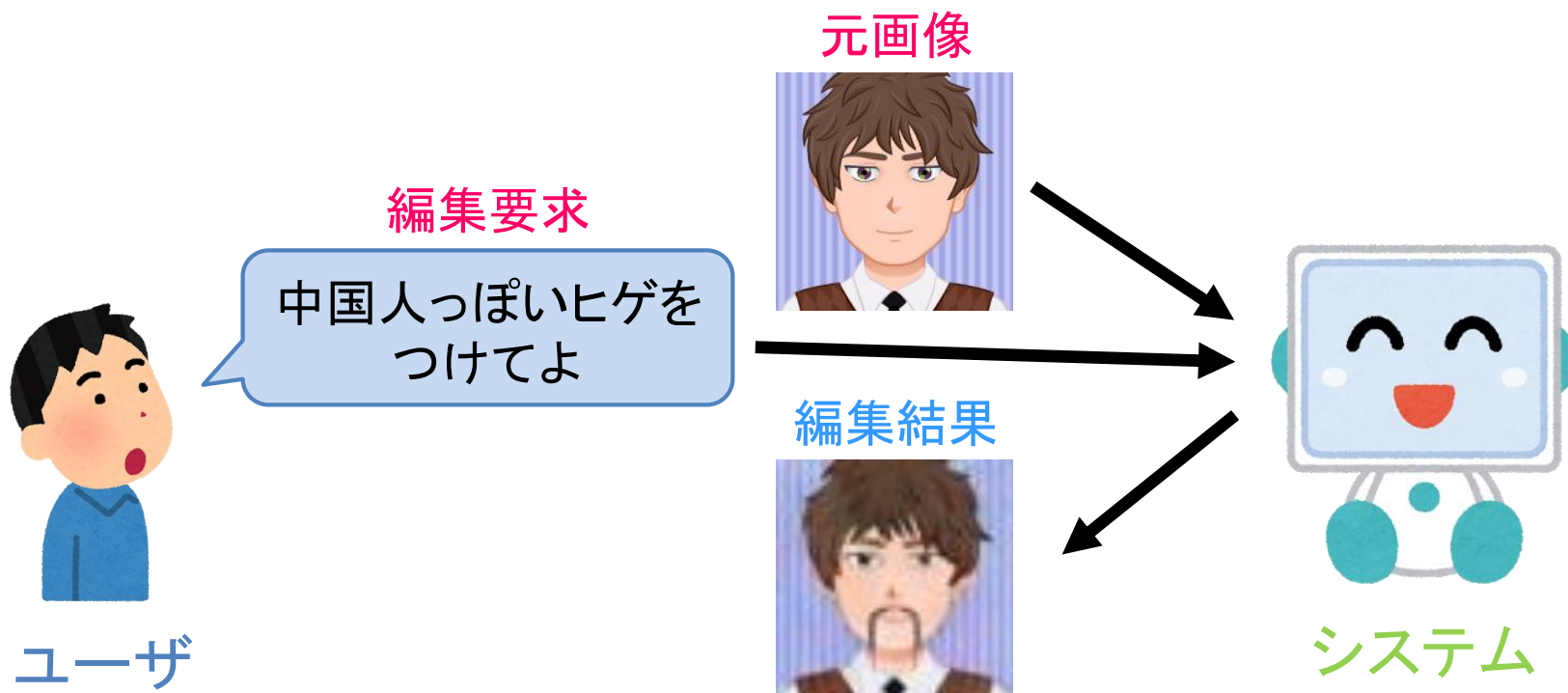
品川政太郎, 吉野幸一郎, Seyed Hossein Alavi,
Kallirroï Georgila, David Traum, Sakriani Sakti, 中村哲

奈良先端大, 理研AIP, USC ICT

NAIST

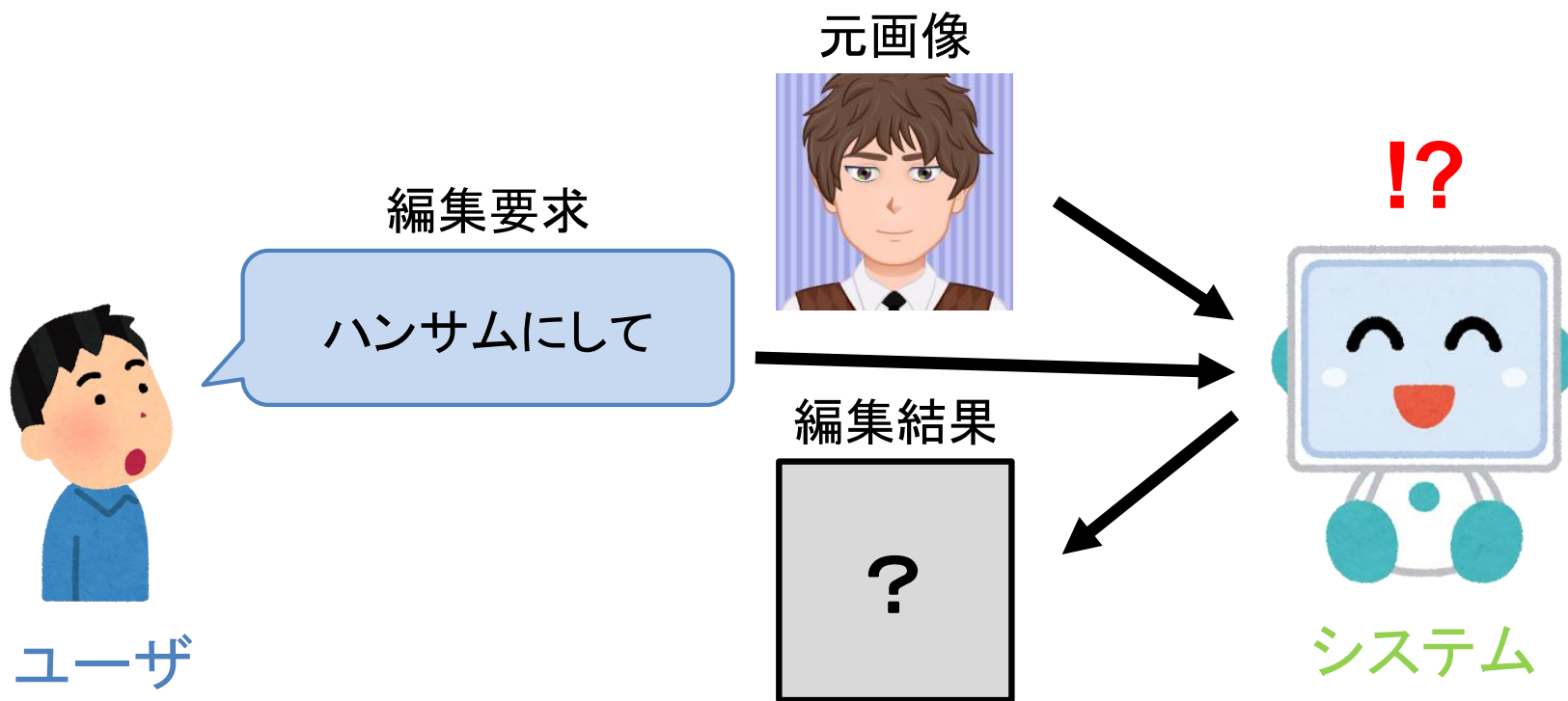
自然言語による画像編集とは？

ユーザが元画像と自然言語による編集要求を入力して
システムが要求に沿った編集結果を出力するタスク



何が難しい？

編集要求(自然言語)の多様性に対応するのが難しい

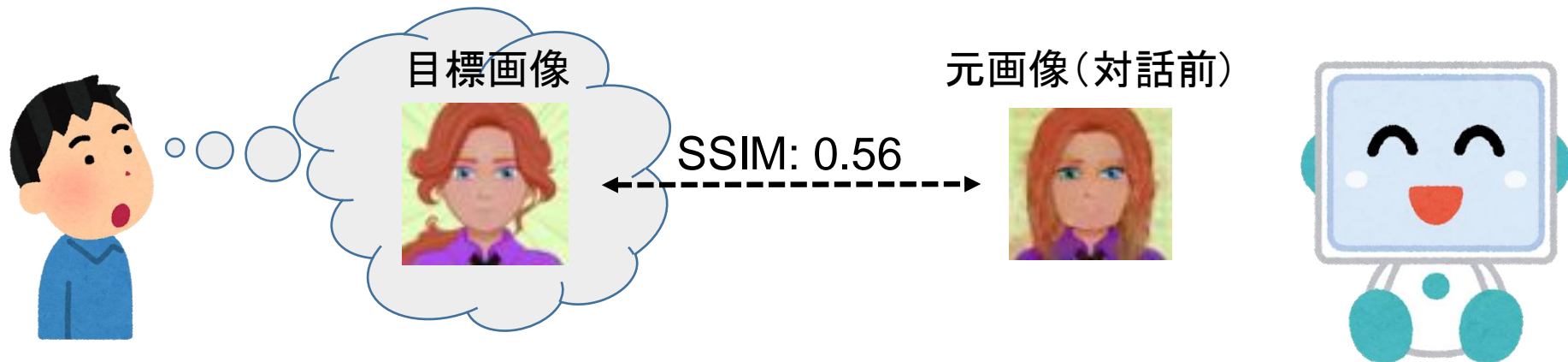


学習した画像編集モデルが苦手とする表現が入ると
おかしい出力を返してしまう

→システムから働きかけて、意図を擦り合わせられれば解決可能

確認戦略に基づく対話的画像編集システム

意図を擦り合わせるため、システムが「確認」を行えるという対話の問題設定として問題を定式化



編集指示文の入力

ポニーテールにして

画像選択

左がいいな

「確認」を選択



元画像の再設定



SSIM: 0.78

どちらが目標を達成するのに適切ですか？

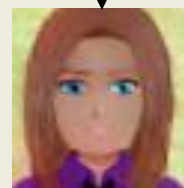
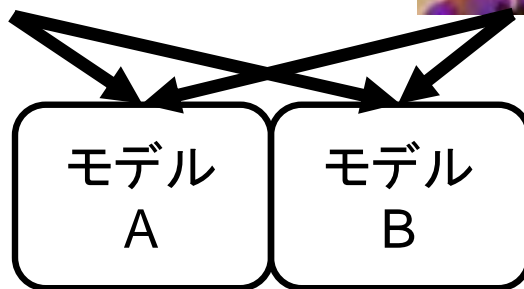
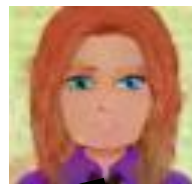
わかりました、他の指示はありますか？

システムの「確認」は何をしている？

1つのモデルによる編集結果が要求に沿えなさそうであれば、異なる特性の複数のモデルの出力から目的に合うものを選んでもらう

編集要求

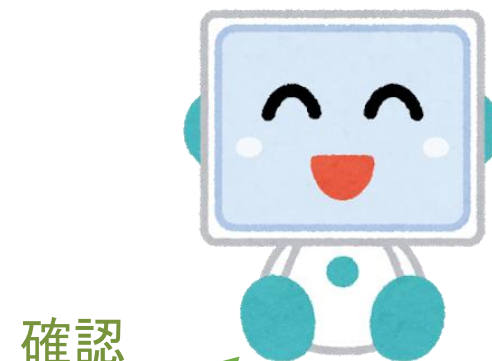
ポニーテールにして



画像選択

左かなあ

両方を提示



確認

どちらが目標を達成するのに適切ですか？

問題：毎回選ぶのはユーザには手間（必要な時だけ確認してほしい）

本研究の貢献

マスクを持つ画像編集モデルを利用して、
マスクのエントロピー計算に基づく確認戦略を提案

結果：無駄な確認が減り、ユーザが画像を選ぶ労力を抑えられた

詳しい結果は下記の論文をご参照ください

**An Interactive Image Editing System Using an
Uncertainty-Based Confirmation Strategy (IEEE Access)**

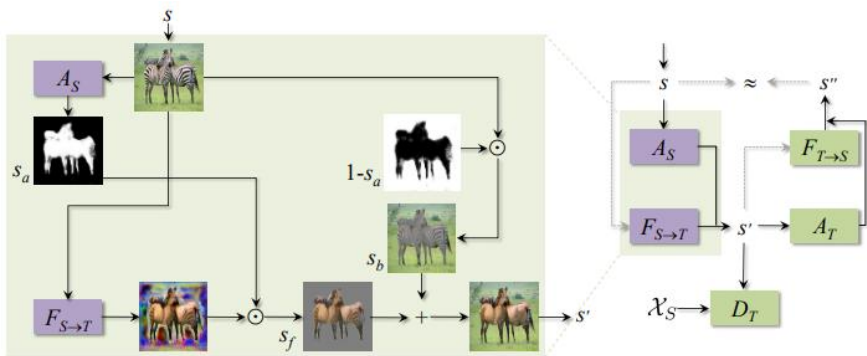
<https://ieeexplore.ieee.org/document/9099288>

確認を行うための基準

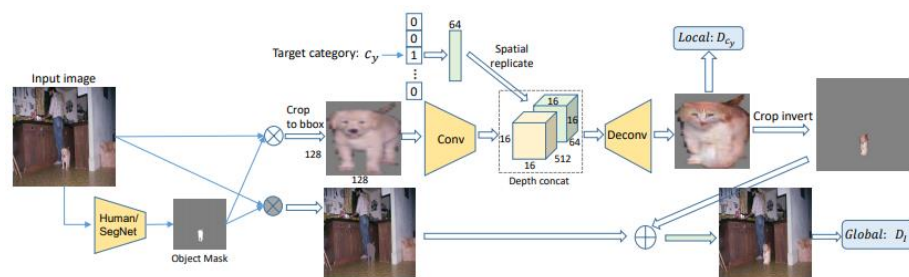
前提: マスクを持つ画像編集モデルに基づいた手法
 確認基準: マスクのエントロピーが閾値より高いと確認する

マスクとは？

- 画像変換モデルに汎用的に使える手法
- 変換したい領域を[0,1]のマスクで指定して制約する



Unsupervised Attention-guided
Image-to-image translation
[Mejjati+, NeurIPS2018]



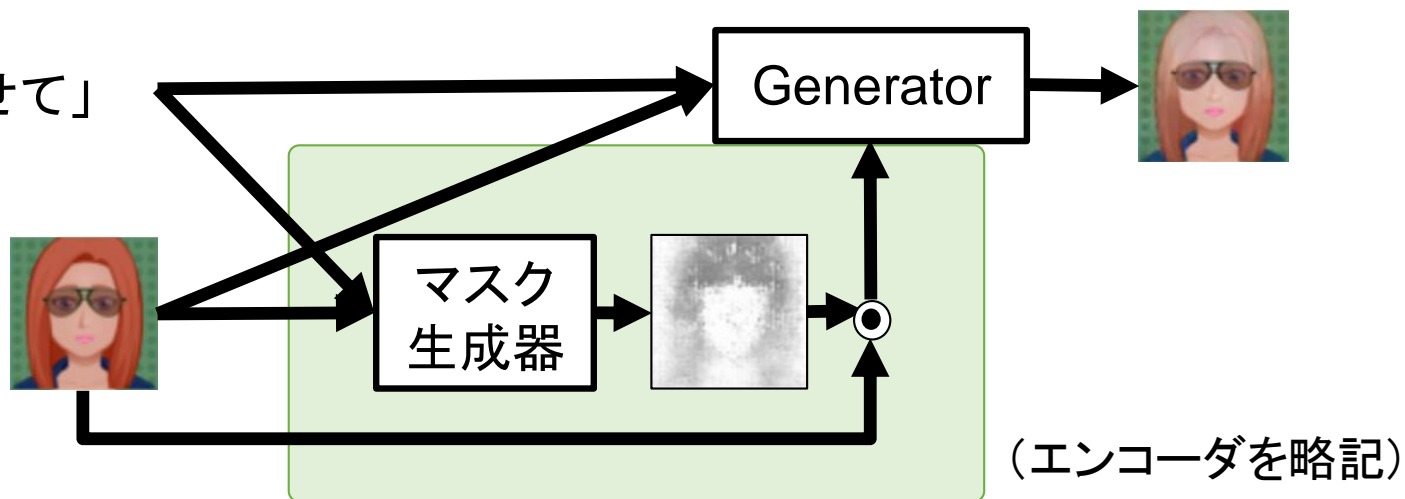
Mask-contrasting GAN
[Liang+, ECCV2018]

確認を行うための基準

自然言語による画像編集におけるマスクありモデル

我々が過去にDCGANベースの手法を提案 [Shinagawa+,MIRU2018]

「頭を禿げさせて」



- 編集要求が意図しない変化を抑制できて全体的な性能が向上
- △髪などの大きい領域の編集が苦手

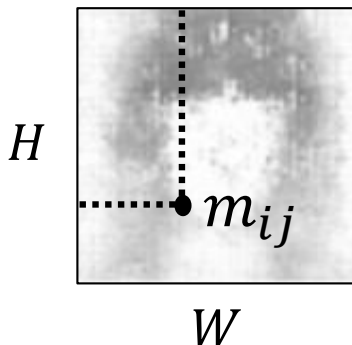
➡ マスクなしモデルとうまく組合わせて使いたい

確認を行うための基準

マスクのエントロピーによる確認基準

これらのマスクはエントロピーを計算できる

- $[0, 1]$ の連続値のときはそのまま
- $\{0, 1\}$ の離散値のときは予測分布から



$$\begin{aligned} & \text{entropy} \\ &= -\frac{1}{WH} \sum_i^H \sum_j^W \{m_{ij} \log(m_{ij}) + (1 - m_{ij}) \log(1 - m_{ij})\} \end{aligned}$$

提案する確認戦略のポイント:

- マスクはどの領域が編集すべき部分なのかの確信度
- エントロピーが閾値を超えたら確信度が低いとして確認を行う

付録：確認戦略に基づく対話的画像編集システムの設定まとめ

使用したデータセット & モデル（詳しくは[Shinagawa+, IEICE2019]）

- AIMIデータセット：アバター顔画像のデータセット
（元画像, 目標画像, 編集指示文）の3つ組データ
- DCGANベースのマスクあり/なし画像編集モデル（一長一短）

モデルの種類	Pros.	Cons.
マスクありモデル	○目や鼻などの小領域の編集が得意	△髪などの大領域の編集ができない
マスクなしモデル	○髪などの大領域の編集もできる	△編集したくない領域まで一緒に編集してしまう

エントロピーに基づく確認戦略（本研究での提案）

マスクは変更すべき領域を二値で表現している

→マスクのエントロピーが高いほどマスクありモデルで失敗する可能性が高いので確認するようにする