Wasserstein Autoencoder を用いた画像スタイル変換

中田秀基、麻生英樹

産業技術総合研究所人工知能研究センター

この成果の一部は、国立研究開発法人新エネルギー・産業技術総合開発機構（ＮＥＤＯ）の委託業務の結果得られたものです。

本研究はJSPS科研費 JP16K00116の助成を受けたものです。

背景

- 画像スタイル変換
  - 入力：コンテント画像とスタイル画像
  - 出力：スタイル画像の「スタイルで」描画されたコンテント画像の「コンテント」
- 古くから研究されていたが近年急速に進展

スタイル　　コンテント

画像は[Gatys, et.al ‘16]より

目的と成果

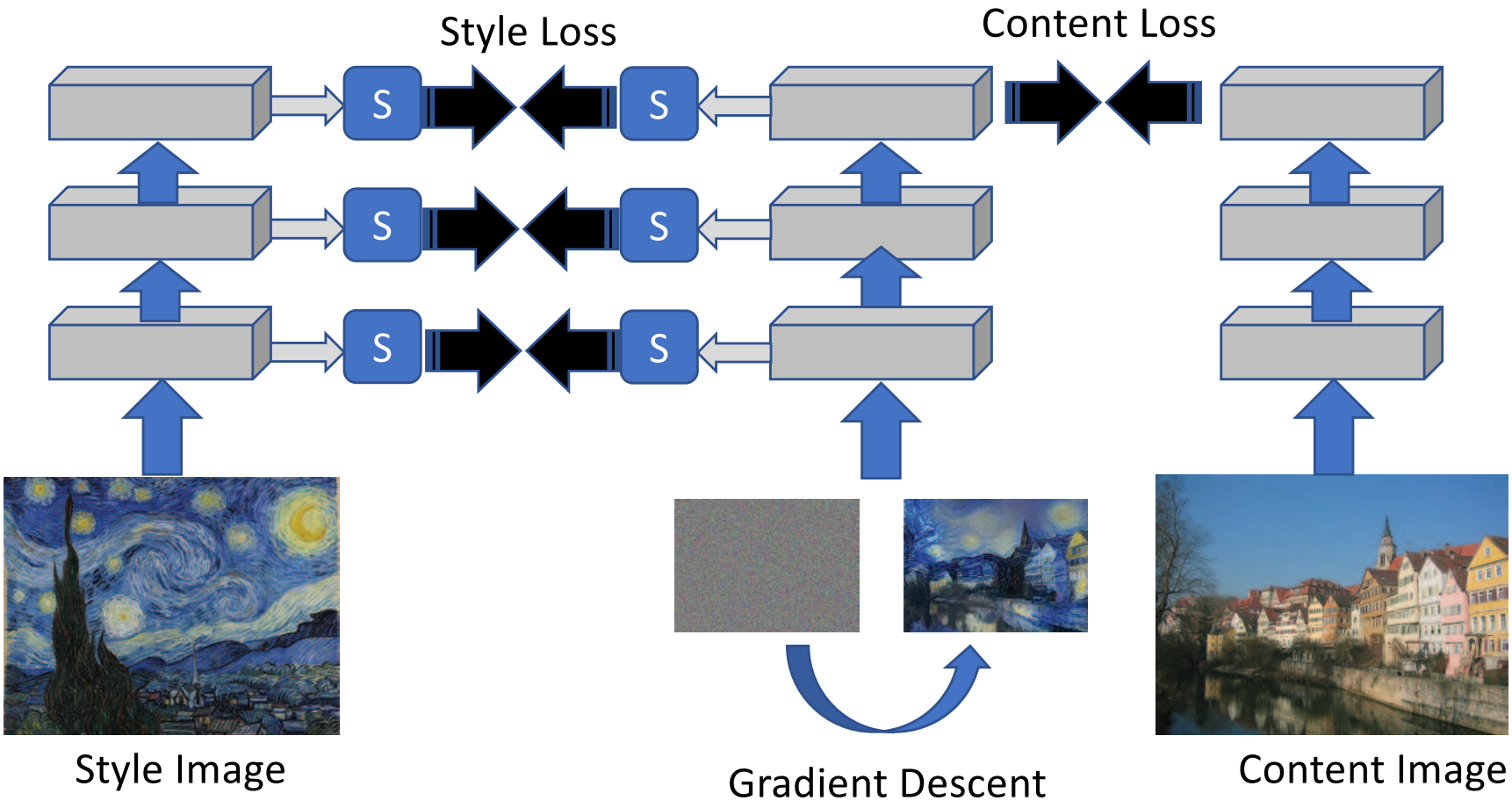
- 既存手法の問題点
  - 画像の生成、もしくはネットワーク訓練に時間がかかる
- 研究の目的
  - 任意のコンテント、スタイルに対して即座に
    スタイル変換を行うことのできる手法の確立
- 先行研究
  - VAEをベースとした手法を提案[PRMU'18]
  - 品質はいまひとつ
- 本研究の成果
  - WAEをベースとした手法を提案
  - 比較的高品質の画像変換が可能であることを確認

発表の概要

- 背景
  - 既存画像変換手法
  - VAE（変分オートエンコーダ）,WAE
- 提案手法
- 評価
- 関連研究
- おわりに
  - まとめと今後の課題

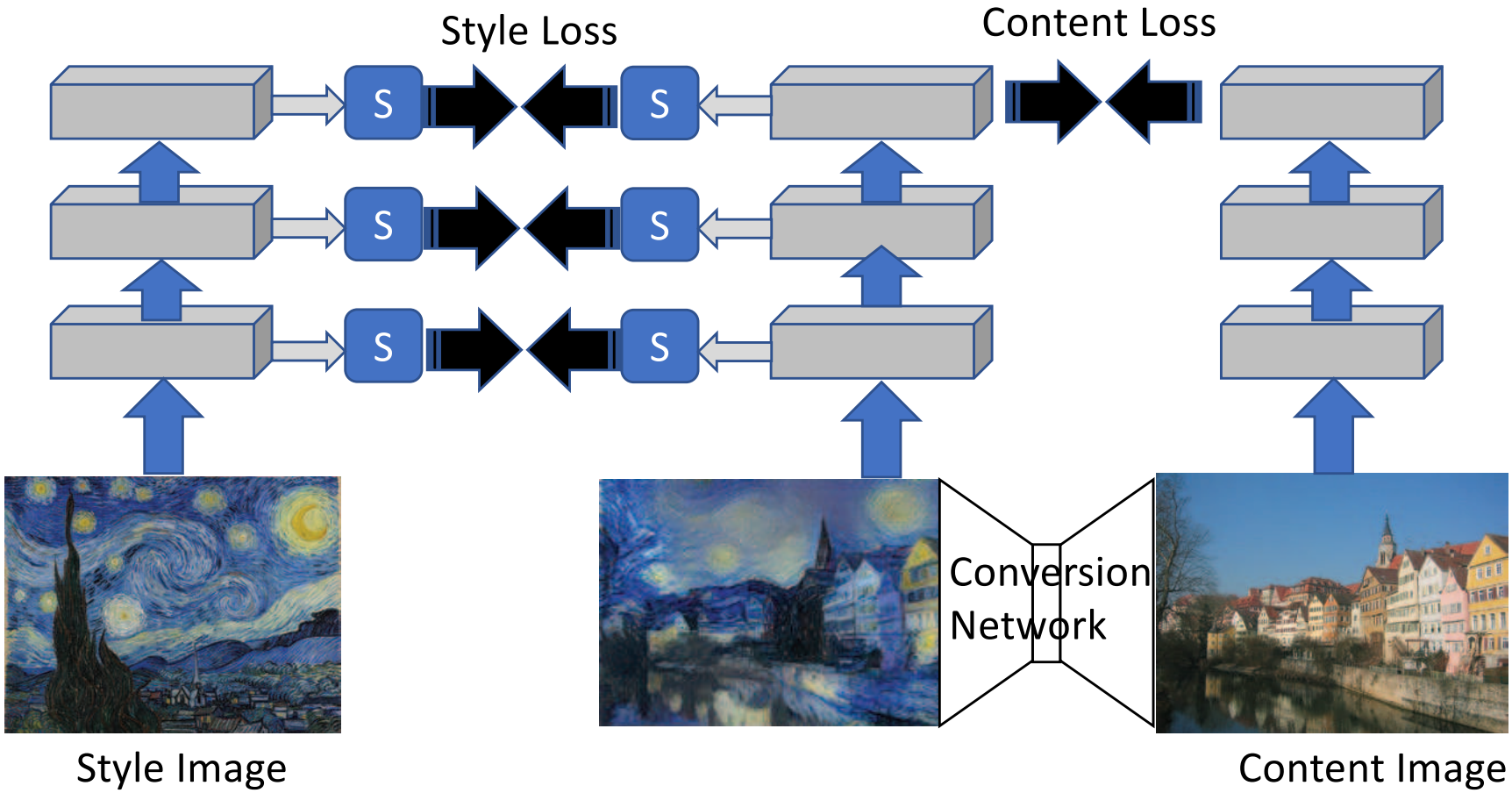
既存画像スタイル変換手法

- ノイズから画像を生成 [Gatys' et.al, '16]
  - 時間がかかる

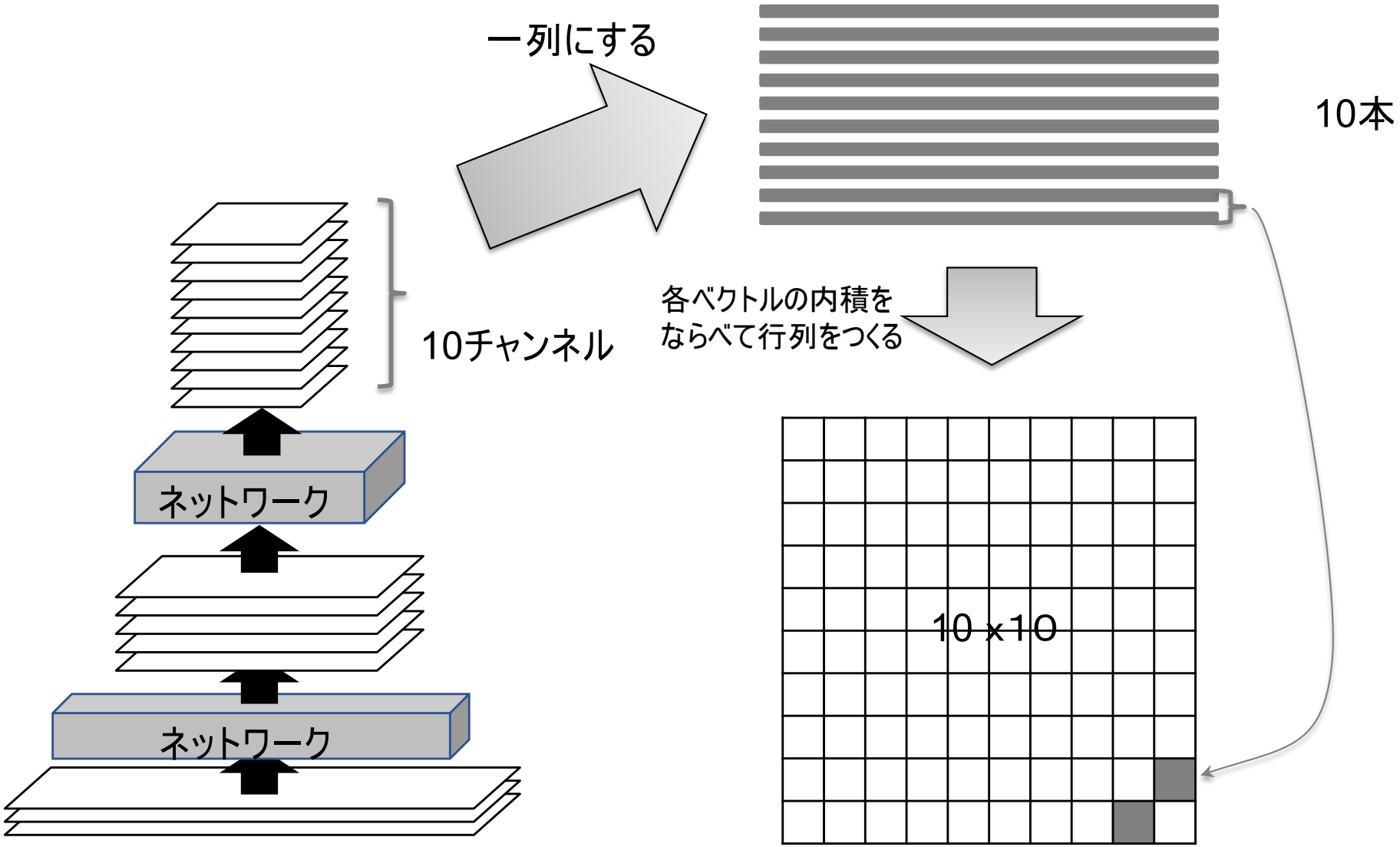


既存画像スタイル変換手法（２）

- 変換ネットワークをトレーニング [Johnson ‘17]
  - 変換自体は高速だがトレーニングはコストがかかる



スタイル行列とは


一列にする
10本
10チャンネル
各ベクトルの内積を
ならべて行列をつくる
ネットワーク
ネットワーク
10 x 10

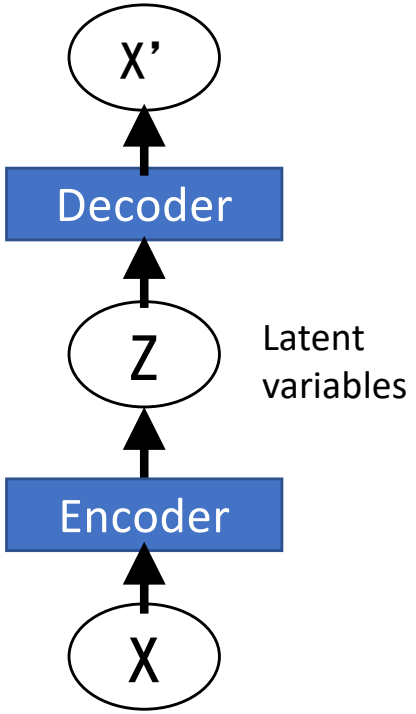
既存手法の問題点

- 時間がかかる
  - Gatysら：画像そのものを最適化
    - コンテント、スタイルどちらを変更しても最適化をやりなおし
  - Johnsonら：特定のスタイルにネットワークを最適化
    - 同じスタイルであれば一つのネットワークで高速に処理できる
    - 別のスタイルにするにはネットワークを訓練し直す必要がある

→ 任意のコンテント、任意のスタイルで即時変換

Variational AutoEncoder

- 元データXから隠れ変数Zを経由して X‘を再現する
  - エンコーダとデコーダの組を同時に訓練する
- 古典的なAutoEncoderとの差異
  - 隠れ変数Zが平均0、分散1の正規分布に近くなるように訓練
  - 隠れ変数の分布と画像の分布が対応
    - 隠れ変数をサンプリングすることで‘意味’のある画像を出力



Wasserstein AutoEncoder

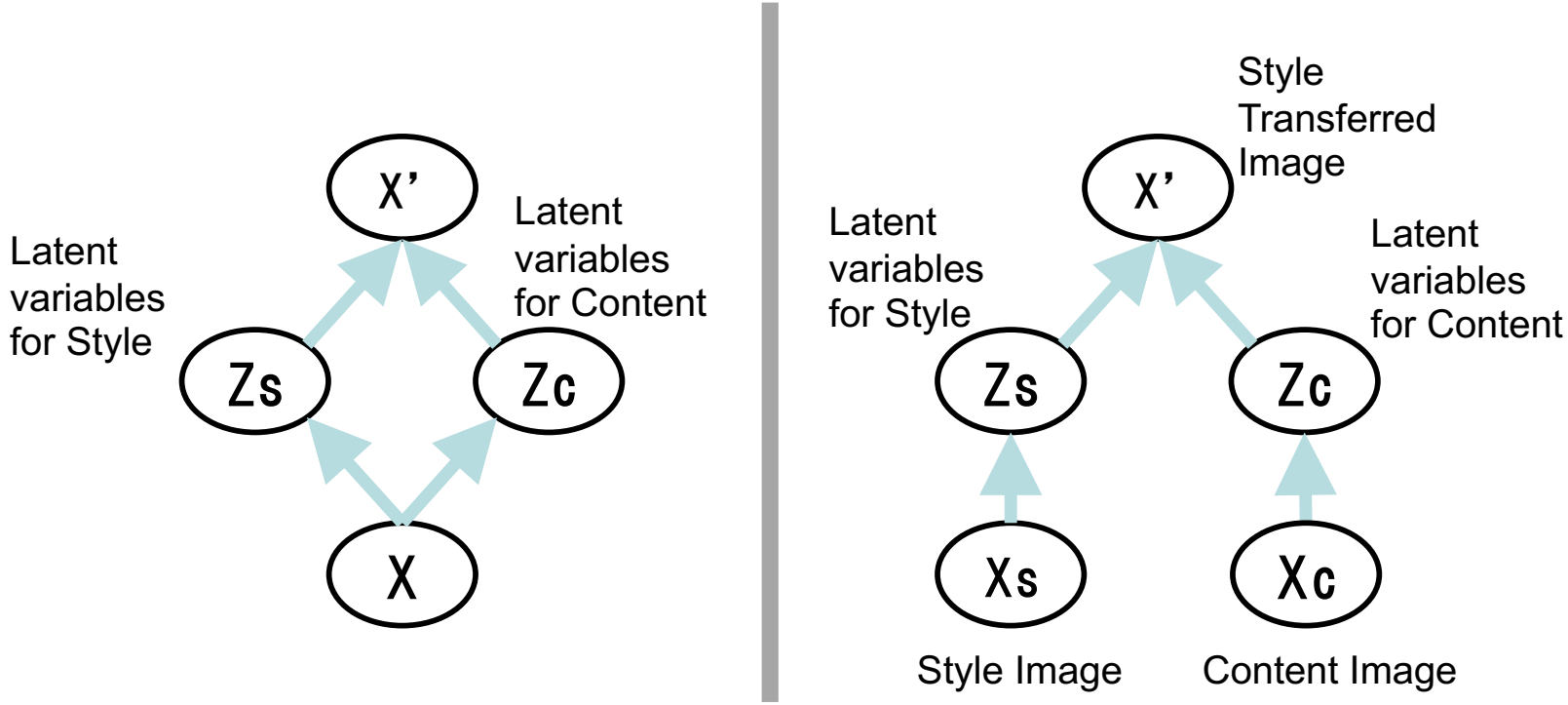
- VAEの亜種
- 隠れ変数zを正規分布に近づけるさいに、KLダイバージェンス最小化ではなく、Wasserstein距離最小化を用いる
- 分布全体の性質をよりよく保存する

発表の概要

- 背景
  - 既存画像変換手法
  - VAE（変分オートエンコーダ）,WAE
- 提案手法
- 評価
- 関連研究
- おわりに
  - まとめと今後の課題

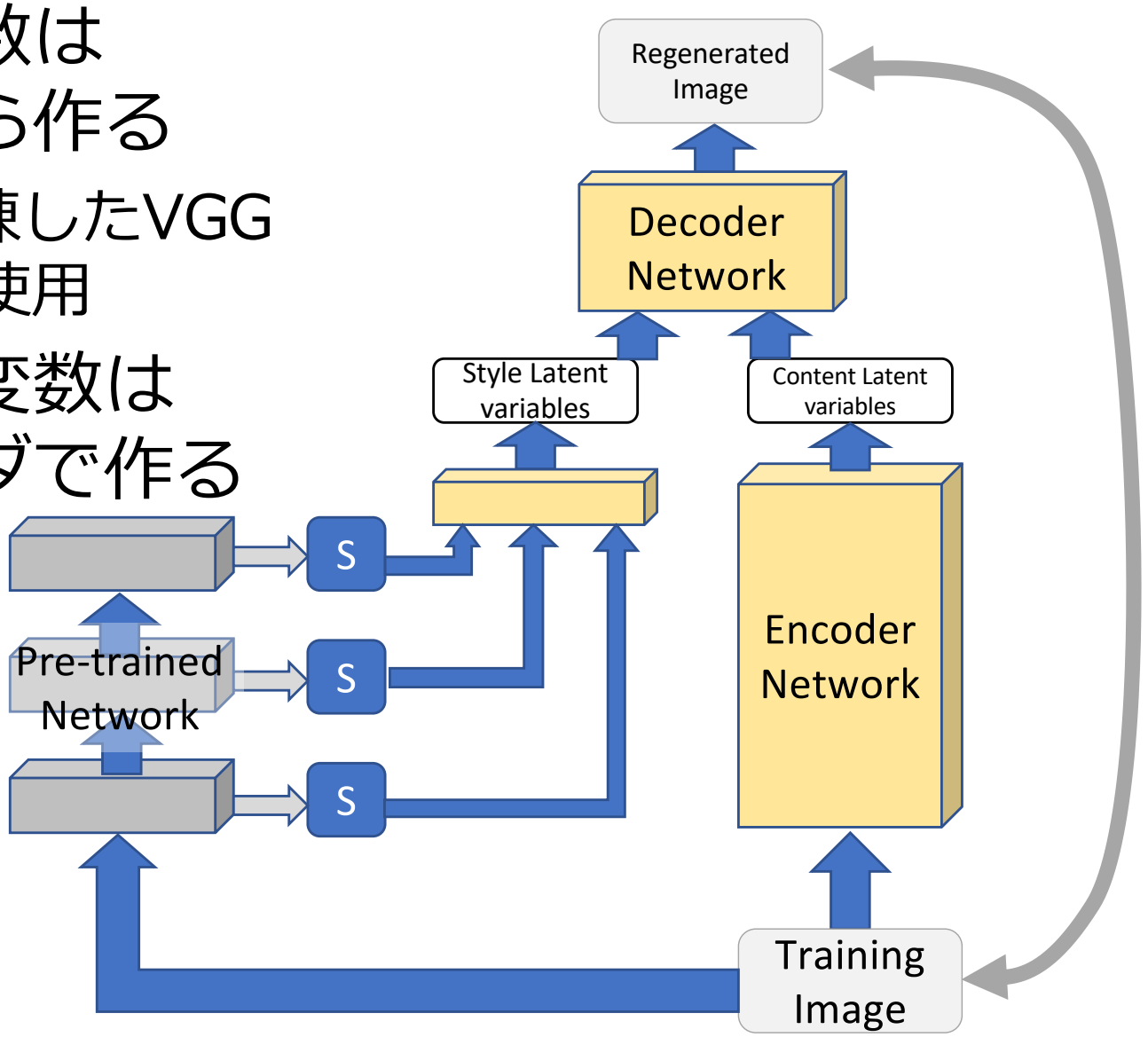
提案手法

- 隠れ変数を、スタイルとコンテントに分離
  - 別の画像から得た隠れ変数を用いればスタイル変換が実現できる



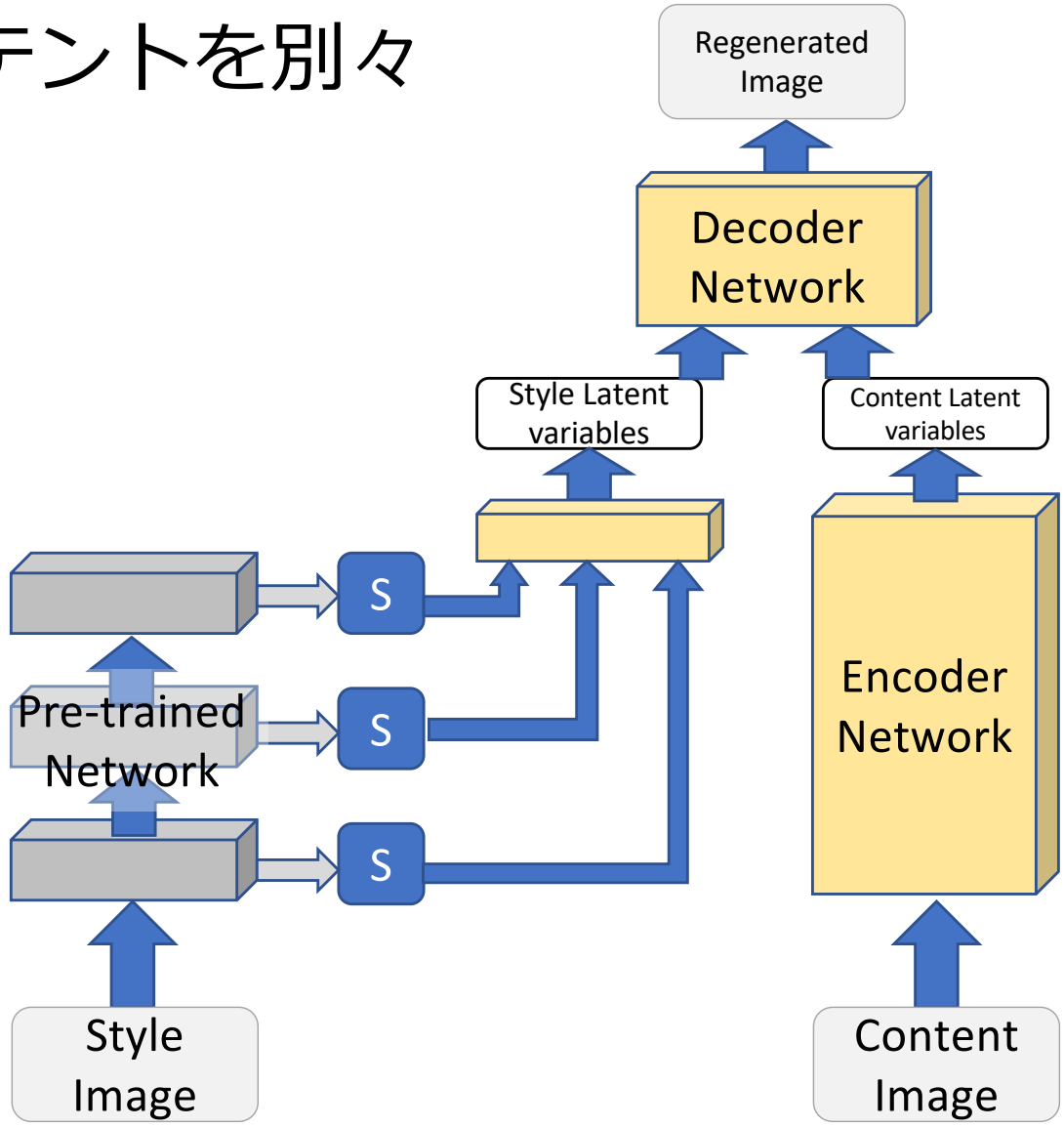
提案ネットワーク

- スタイル隠れ変数は
  スタイル行列から作る
  - ImageNetで訓練したVGG
    ネットワークを使用
- コンテント隠れ変数は
  通常のエンコーダで作る



スタイル変換画像生成時

- スタイルとコンテントを別々に入力



発表の概要

- 背景
  - 既存画像変換手法
  - VAE（変分オートエンコーダ） ,WAE
- 提案手法
- 評価
- 関連研究
- おわりに
  - まとめと今後の課題

評価

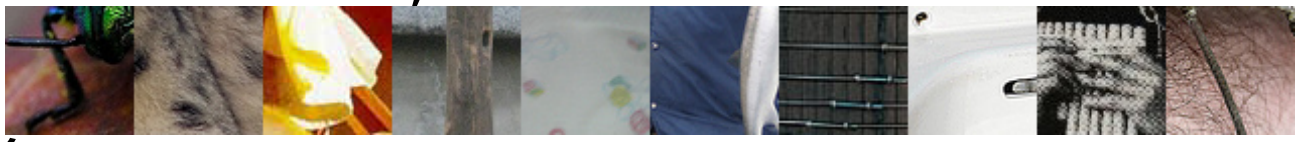
- 評価方法
  - 画像再構成
    - 基本的なエンコーダ・デコーダの性能を評価
  - CelebA画像をスタイル変換

- 評価ポイント
  - 訓練時に用いるデータセットの多様性の影響
  - コンテントとスタイルの寄与比率による影響

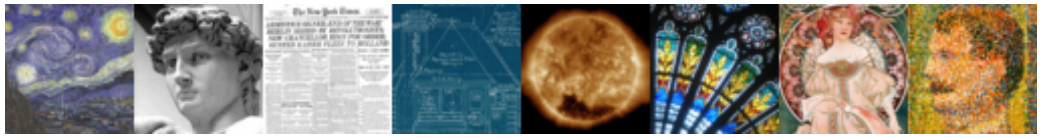
データセット

- CelebA
  - 顔を中心にクロップ、193,800枚
- アニメ顔画像セット
  - リサイズ、14,490枚

- ImageNet
  - 中心部をクロップ、196,371枚



- スタイル画像



実験設定

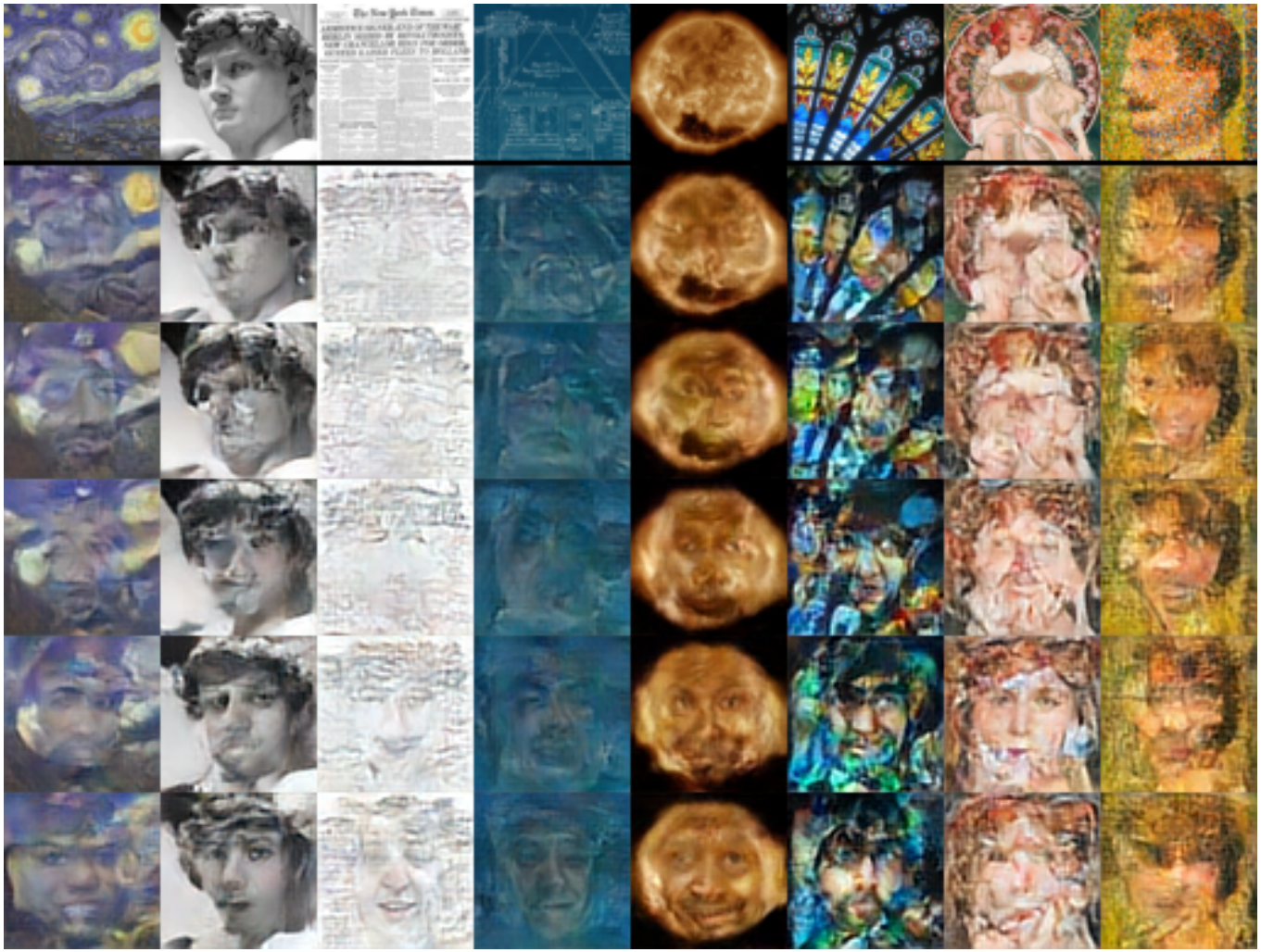
- 学習画像多様性の貢献
  - CelebAのみで学習
  - CelebA＋アニメ＋Imagenetで学習

- コンテント隠れ変数の数でコンテントの寄与を調整
  - 変数の数：512,256,128,64,32

再構成 ClebAのみで学習

再構成
ClebA+Anime+ImageNetで学習

再構成 ClebAのみで学習



再構成
ClebA+Anime+ImageNetで学習

オリジナル

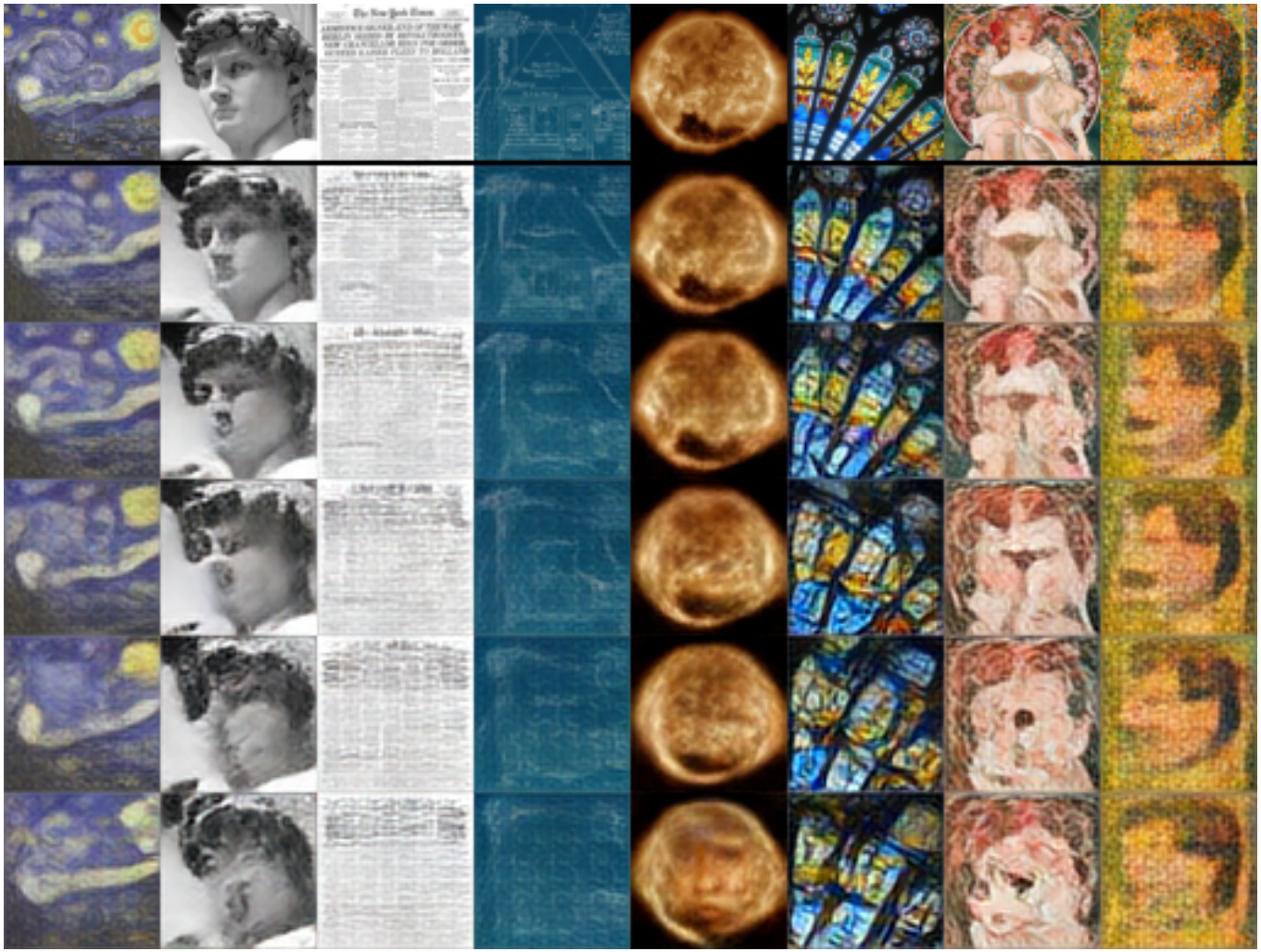
コンテント=512

コンテント=256

コンテント=128

コンテント=64

コンテント=32



スタイル変換、ClebAのみで学習

コンテント

スタイル

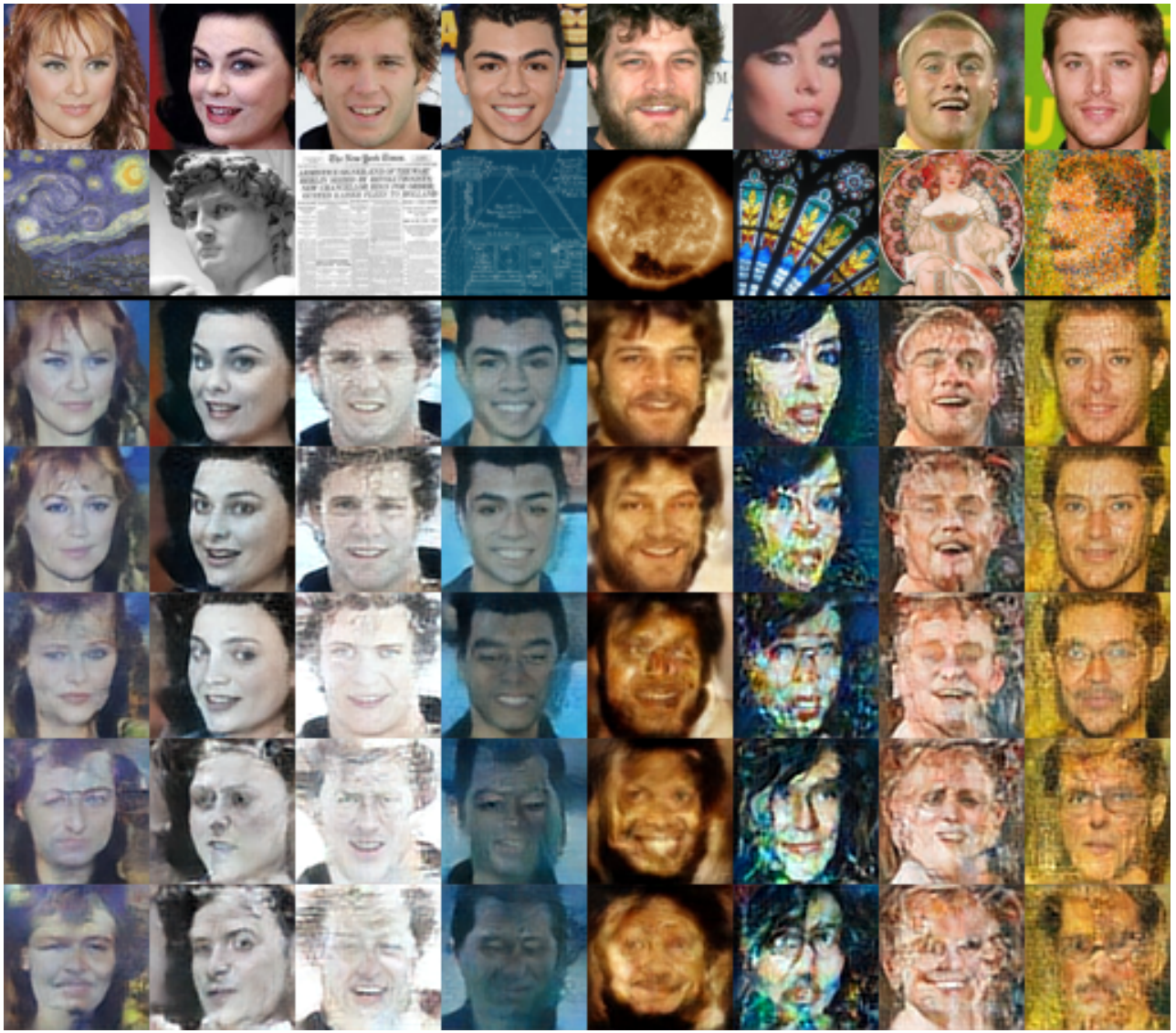
コンテント=512

コンテント=256

コンテント=128

コンテント=64

コンテント=32



スタイル変換、ClebA+Anime+ImageNetで学習

コンテント

スタイル

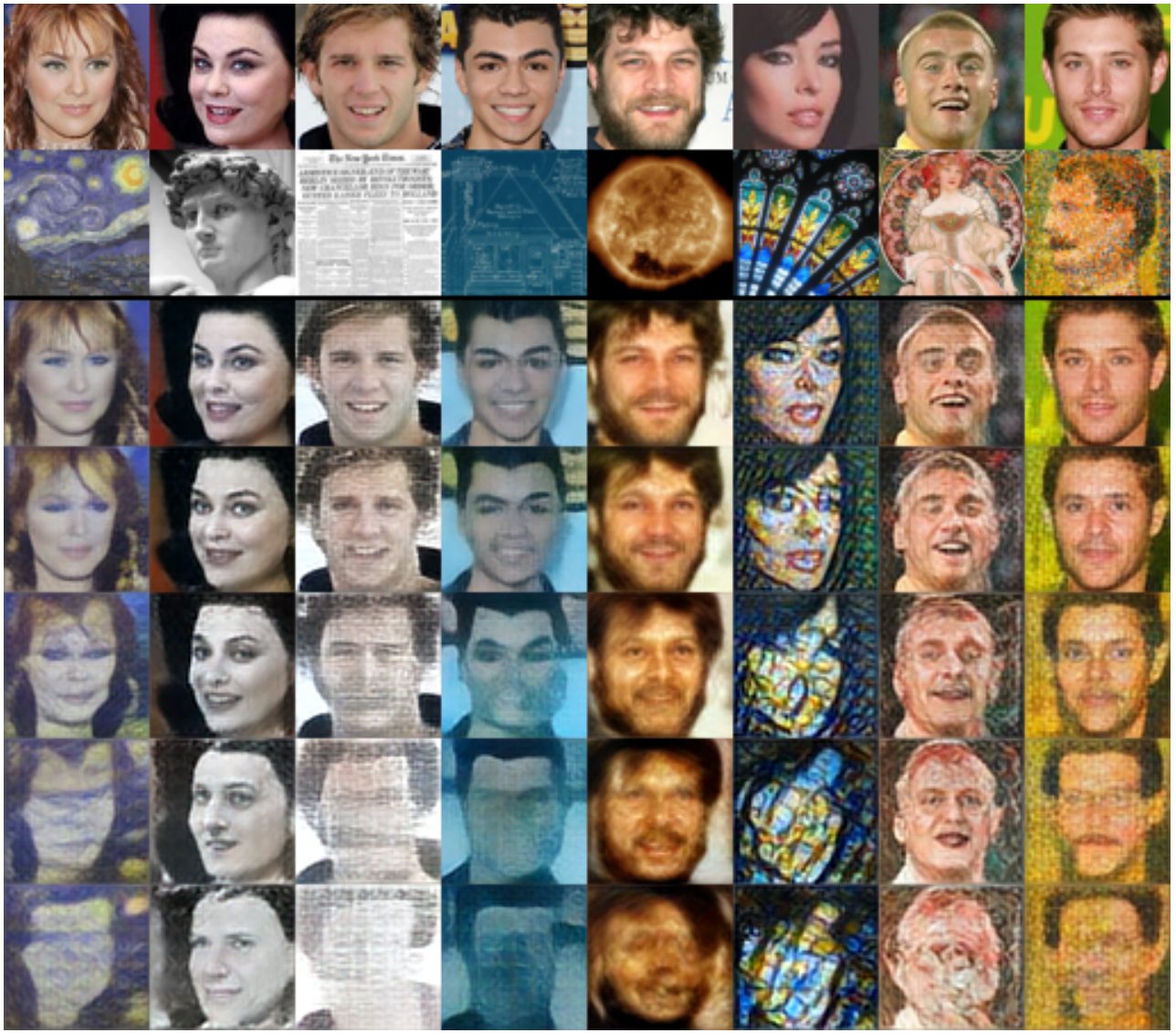
コンテント=512

コンテント=256

コンテント=128

コンテント=64

コンテント=32



AIST

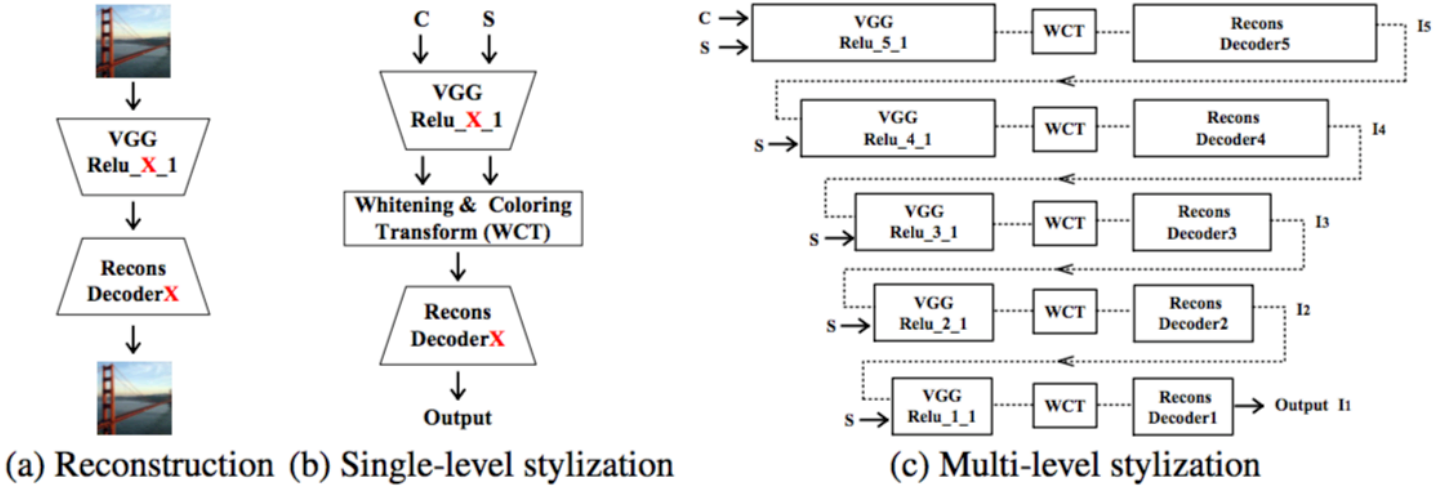
- 混合画像
- 隠れ変数128

発表の概要

- 背景
  - 既存画像変換手法
  - VAE（変分オートエンコーダ） ,WAE
- 提案手法
- 評価
- 関連研究
- おわりに
  - まとめと今後の課題

関連研究

- ホワイトニング、カラーリングによる手法
  - [Li, et.al, NIPS ’17, ArXiv 182.0647]



(a) Reconstruction (b) Single-level stylization (c) Multi-level stylization

発表の概要

- 背景
  - 既存画像変換手法
  - VAE（変分オートエンコーダ） ,WAE
- 提案手法
- 評価
- 関連研究
- おわりに
  - まとめと今後の課題

おわりに

- まとめ
  - WAEをベースとしたスタイル変換手法を提案
    - 任意のスタイル、コンテントに対して即時に画像を生成可能
  - 一定の品質が得られた
- 今後の課題
  - 生成ネットワークの表現力
    - GANなどの利用
    - PixelCNNなどのより強力なデコーダの利用