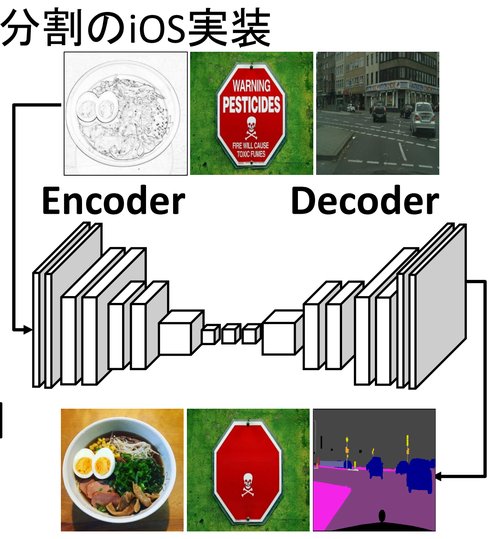
DS-12 ConvDeconvNetの効率的モバイル実装による画像変換・物体検出・領域分割リアルタイムiOSアプリ群

丹野 良介，泉 裕貴，柳井 啓司(電気通信大学)

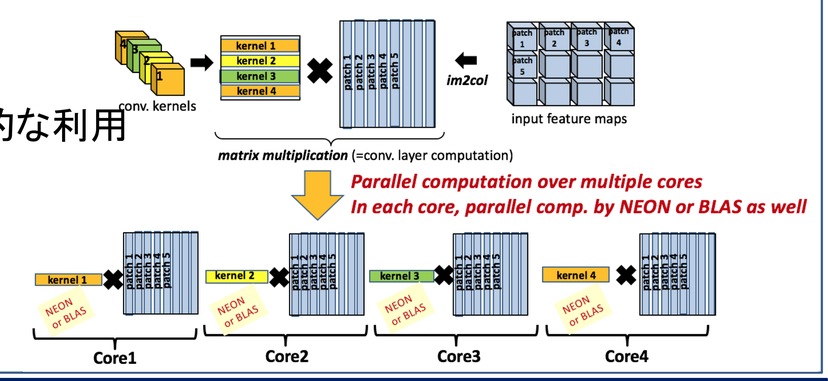
1. はじめに

- ConvDeconvNetによる画像生成，領域分割のiOS実装
  - 画像変換をpair画像で学習pix2pix[2]
    - 線画着色，領域分割
  - pix2pixをUnpairに拡張CycleGAN[3]
    - 素材クラスの変換(マテリアル変換)
  - Multi Style Transfer[4]
    - [1]のネットワークの拡張，単一モデルで複数のスタイルを任意重みで合成
  - 九州大学内田研とのコラボ研究[6, 7]
    - 情景画像内の文字隠蔽ネットワーク
      - **PS2-22で中身のポスター発表中！**



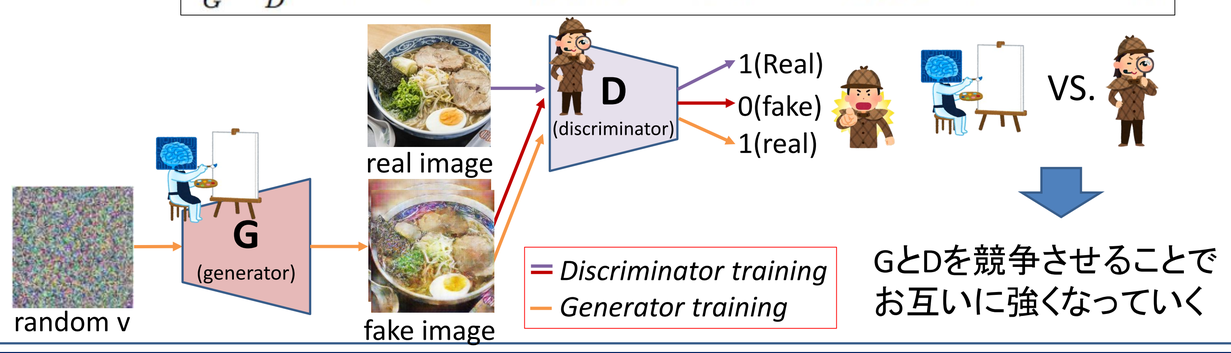
2. モバイル実装ワークフロー

- Caffe2C / Chainer2C
  - 深層学習フレームワークCaffe/Chainerで学習したネットワークをiOS/Androidで動作可能なC言語コードに自動変換
  - オリジナルの順伝搬演算ライブラリとリンク → 高速動作可能
- マルチスレッド
  - BlasLibrary (iOS)
  - Neon (Android)の効率的な利用
- 高速化の工夫
  - 事前計算
  - Sparse Relu
  - Subpixel Conv (Pixel Shuffler)

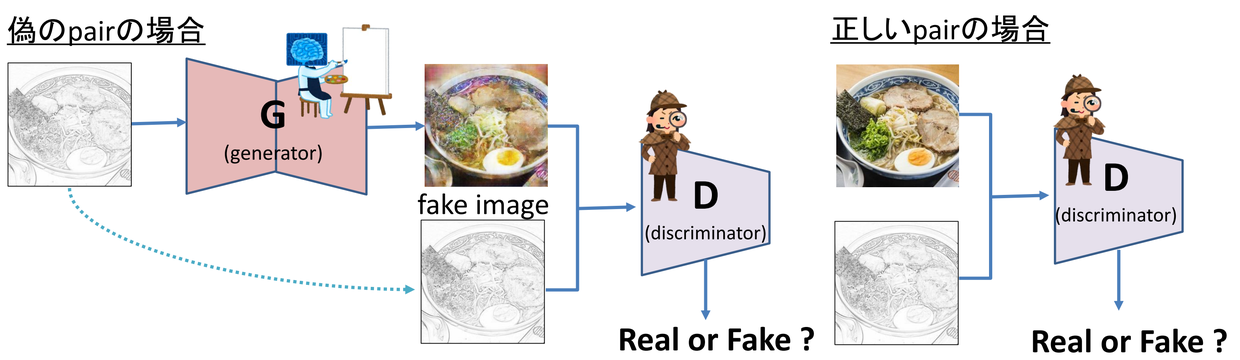


3. pix2pix[2]

- Generative Adversarial Network(GAN)
  - 用意したサンプルデータ群に似たデータを生成
    - G: Generator(生成器): ランダム生成ベクトルを入力，サンプル群に似たデータを出力することで，Dを騙すことを目的
    - D: Discriminator(識別器): 受取ったデータが，サンプル群のデータ(real)か，Gにより生成されたデータ(fake) か識別することを目的
- GANは，以下の価値関数で表現されるミニマックスゲームとして定式化: $\min_G \max_D V(D,G) = \mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}(\boldsymbol{x})}[\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z})))]$

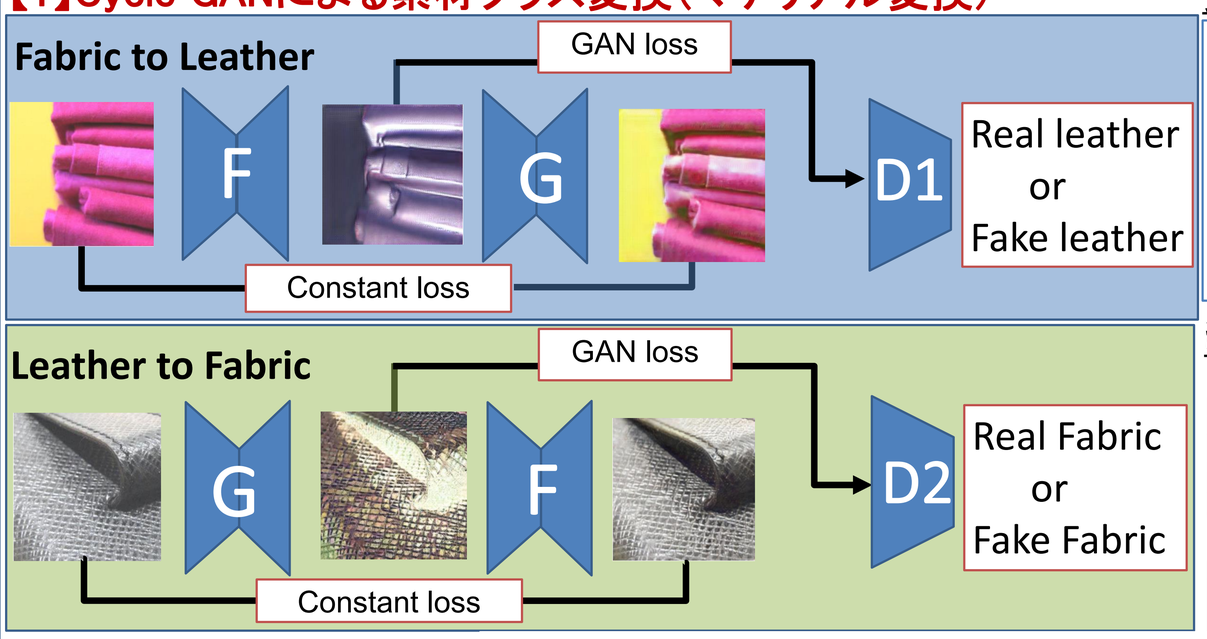


- pix2pix
  - **pair画像間の違いを学習して，その差を補う形で画像を出力**
    - 従来，似た問題設定でも個別のモデルで扱われていた
      -> 似たタスクを1つのモデルで変換できるようにする
      ex. 線画着色，領域分割，カラー化…
    - Generatorにノイズではなく，変換前の画像も入力
    - Discriminatorに損失関数そのものを学習



4. CycleGAN[3]

【1】Cycle GANによる素材クラス変換（マテリアル変換）

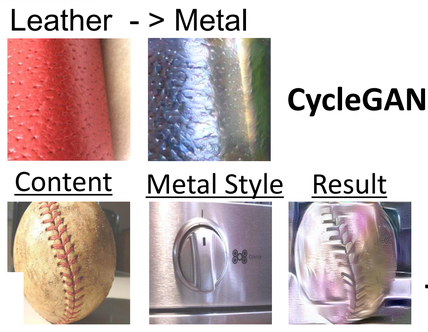


元に戻せる範囲内で変換　　F: fabric to leather　G: leather to fabric

実験

- Train
  - UFMD(岡谷ら)
  - 1000 images *2class
- Test
  - FMD

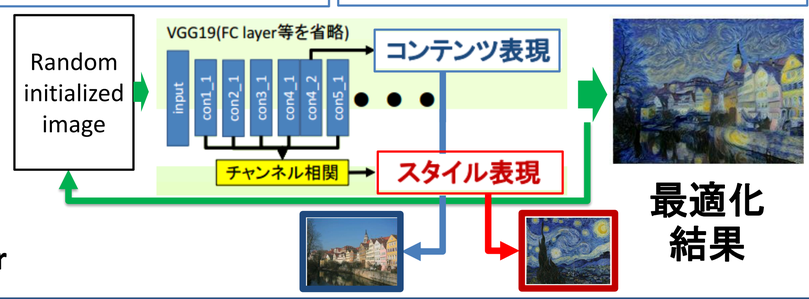
結果



【2】Cycle GANとStyle Transferの比較

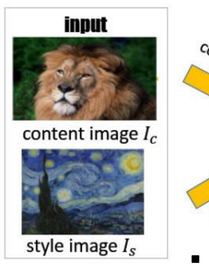
CycleGAN
- 学習
  - 画像集合のペア
- 構造の維持
  - 復元可能な範囲
- 素材の変換
  - D1GAN

Style Transfer
- 最適化
  - 画像ペア
- 構造の維持
  - Feature mapの類似度
- 素材の変換
  - Style vector の類似度

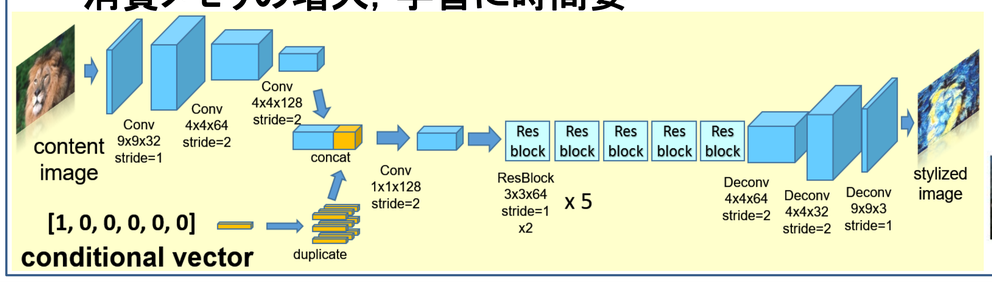


5. Multi Style Transfer[4]

- Neural Style Transfer(Gatys et al, 2015)
  - 画像の形状を保持したまま画風を変換
  - **順伝搬及び誤差逆伝播を複数回**繰り返す
  - GPUで**画像の生成に数十秒程度**
- Fast Style Transfer[1]
  - 特定の画風変換を準伝搬で行う CNN を学習
  - 非常に高速に画像を変換可能
  - **1 つのモデルで単一の画風変換のみ学習**
  - **消費メモリの増大，学習に時間要**

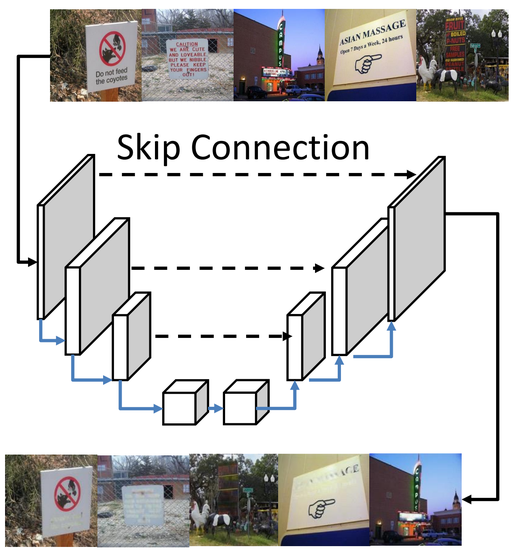


- **[1]の拡張**
- **単一モデルで複数のスタイルを任意重みで合成**



6. 九州大学内田研とのコラボ研究[6, 7]

- 情景画像内の文字を隠蔽するネットワークのモバイル実装担当
  - 文字有画像を入力すると，文字部分が隠蔽されて出力されるよう学習
- 学習データ
  - Train: Flickr 3019枚+ ICDAR 229枚
  - Test: ICDAR 233枚
- CNN構造
  - U-Netを利用
    - 画像サイズが同じものを深い層から段階的に統合
    - 局所的特徴を保持したまま全体的位置情報の復元を可能に
- **PS2-22で中身のポスター発表中！**



参考文献

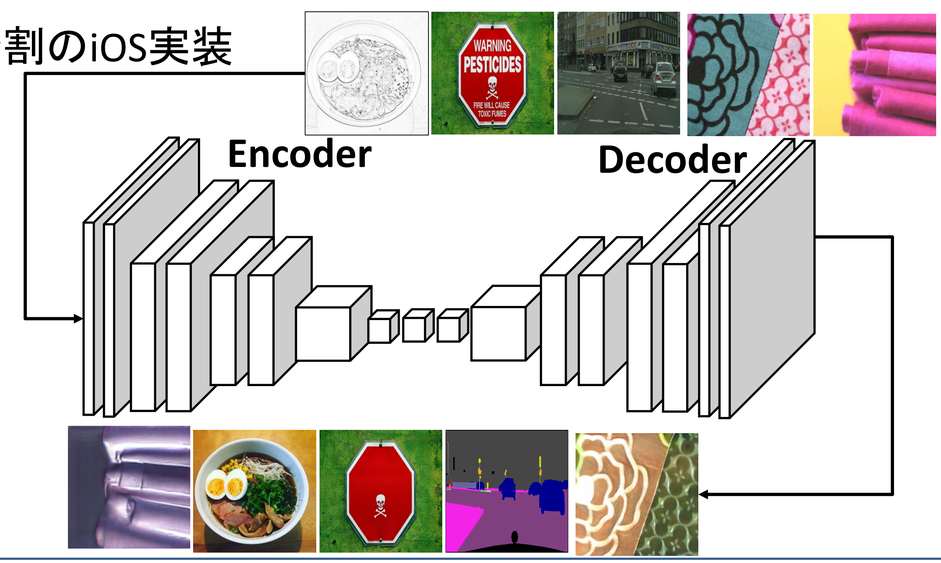
[1] J. Johnson et al.: Perceptual Losses for Real-Time Style Transfer and Super-Resolution, ECCV, 2016.
[2] I. Phillip et al.: Image-to-Image Translation with Conditional Adversarial Networks, CVPR, 2017.
[3] J. Zhu et al.: Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks, ICCV, 2017.
[4] K. Yanai et al.: Conditional Fast Style Transfer Network, ICMR, 2017
[5] K. Yanai et al.: Efficient Mobile Implementation of a CNN-based Object Recognition System, ACM MM, 2016.
[6] T. Nakamura et al.: Scene Text Eraser, ICDAR, 2017.
[7] 中村 俊貴 他: 畳みこみニューラルネットワークを用いた情景画像内の文字隠蔽, MIRU, 2017.

DS-12 ConvDeconvNetの効率的モバイル実装による画像変換・物体検出・領域分割リアルタイムiOSアプリ群

丹野 良介，泉 裕貴，柳井 啓司(電気通信大学)

はじめに

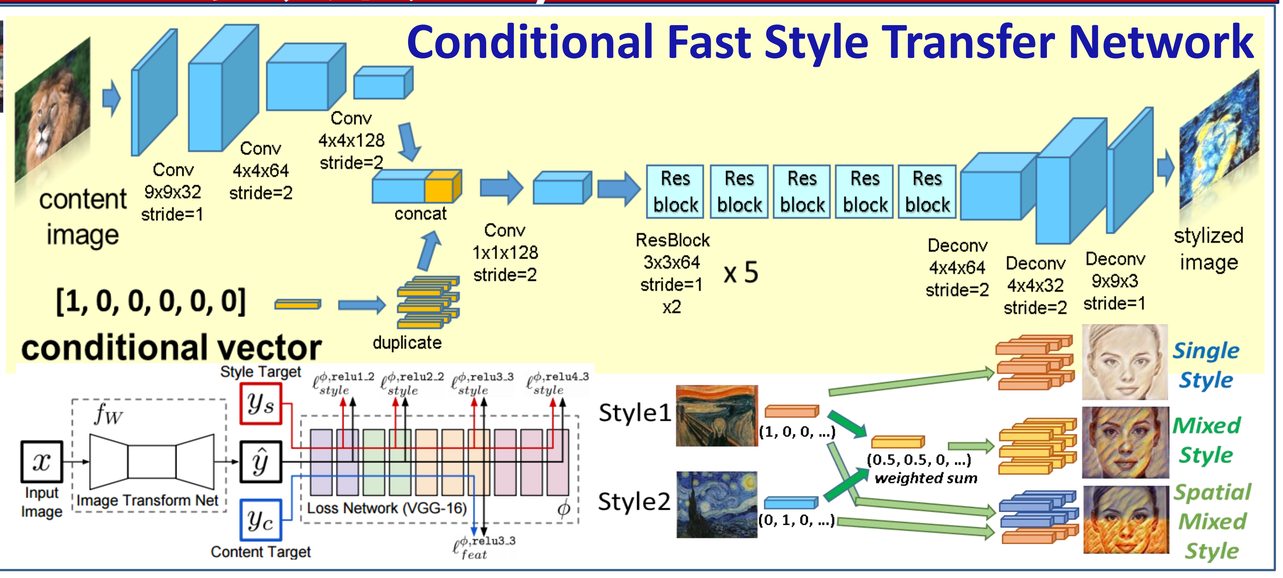
- ConvDeconvNetによる画像生成，領域分割のiOS実装
  - 画像変換をpair画像で学習pix2pix[2]
    - 線画着色，領域分割
  - pix2pixをUnpairに拡張CycleGAN[3]
    - 素材クラスの変換(マテリアル変換)
  - Multi Style Transfer[4]
    - [1]のネットワークの拡張，単一モデルで複数のスタイルを任意重みで合成
  - 九州大学内田研とのコラボ研究[6, 7]
    - 情景画像内の文字隠蔽ネットワーク
      - **PS2-22で中身のポスター発表中！**



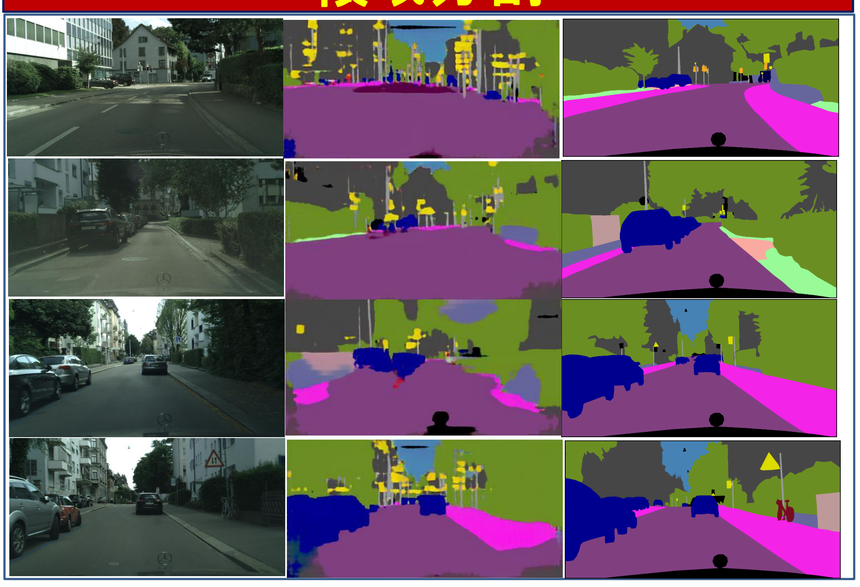
デモ動画

線画着色

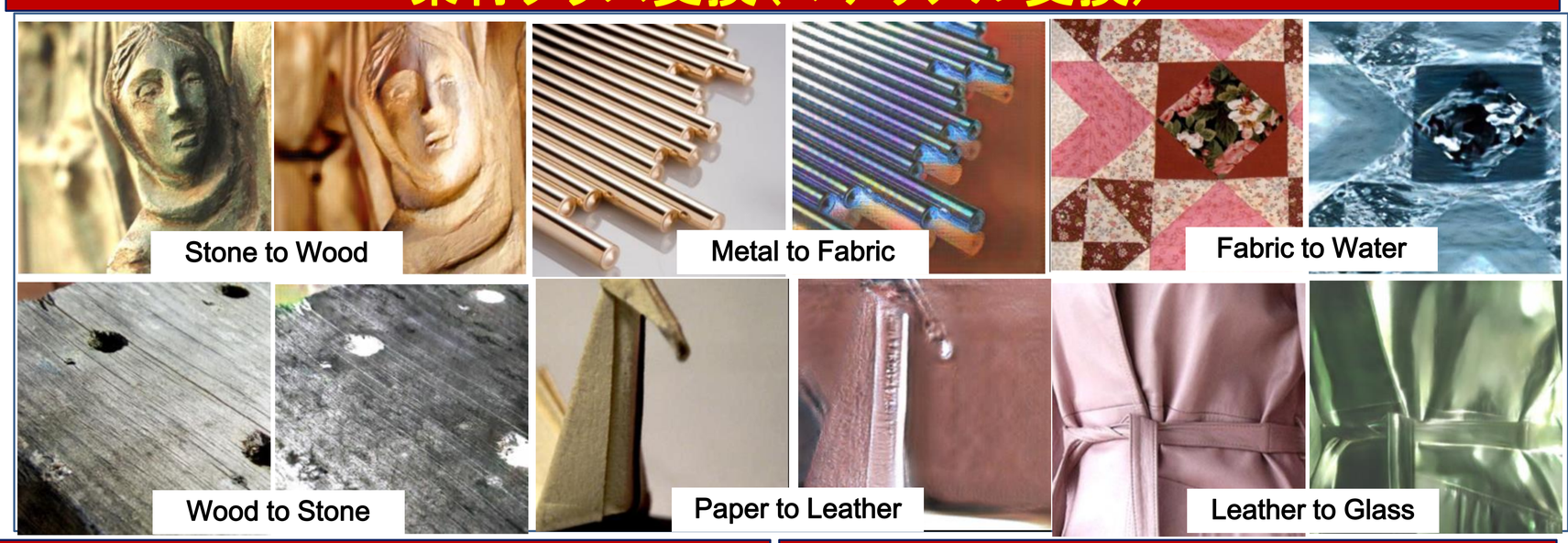
マルチスタイル変換 (PS3-27で発表予定！)



領域分割



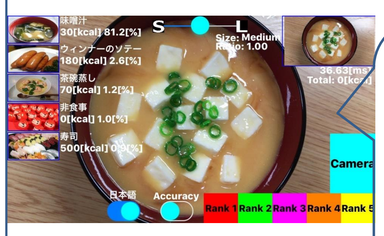
素材クラス変換（マテリアル変換）



文字隠蔽（九大内田研とのコラボ）(PS2-22で発表中！)

PS2-22で中身の発表中！

アプリダウンロード



- 画像認識アプリ(公開済)

DeepFoodCam

- 画像変換アプリ(公開済)

DeepStyleCam

Apple版のみ