A Proof-of-concept use of Generative Adversarial Networks (GANs) to Remove Pen Marks from Whole Slide Images

FangYao Hu¹ and Cleopatra Kozlowski²

¹Dept. of Safety Assessment
²Dept. of Oncology Biomarker Development
Genentech Inc.
Conflict of Interest

- Fang Yao and I are both employees and stock-holders of Genentech / Roche
Outline

- Introduction
  - Why remove marks from slides?
- Methods
  - Preparation of training set samples
  - Generative Adversarial Networks (GANs)
- Results
  - Example images
  - Computing GAN errors
- Future Directions
Introduction

- Why do pathologists mark slides?
  - Tumor area marked for RNA extraction
  - Notes on diagnosis and other metadata
- Why can’t we just clean the slides?
  - Only the images (but not the slides) may be available
  - The marks may be still needed for downstream processes
  - Too much time required to clean
  - Cleaning can occasionally damage slides
  - Marker may be permanent
Benefits of restoring the original image

- Allows other analyses (e.g., image segmentation, patient response prediction, etc.) to be applied without losing information
- Solve anonymization issues (scribbles may have patient identifying information)

- In our hands, color deconvolution was insufficient due to marker
  - Opacity variation
  - Color gradients
- **GANs** might be an easier, faster approach
Generative Adversarial Networks (GANs)

Input Image

Generator

The ‘art forgery expert’

Generated Image

Discriminator

Is it real or generated?

The ‘art critic’

Training Image
Methods

Precious clinical dataset (leave untouched) Hematoxylin and Eosin (H&E) Slides

Build a training dataset of procured samples (100 slides) similar to clinical dataset

1. Scan images
2. Clean set of H&E tissue slides
3. Mark up (at random) With same marker used for clinical dataset (red, green, blue and black)
4. Register clean and marked images
5. Hamamatsu Slide scanner
6. High magnification tile extraction on only marked areas
7. Slidematch software Microdimensions

Hamamatsu Slide scanner

Slidematch software Microdimensions
Identify marker areas, then only add data from marked area for GAN training

- 86 slides (~87,000 tiles) for training
- 10 slides for validation
- 3 slides for testing
- 1 failed registration step
Example of Paired Tiles used for Training

With Marks

Original (unmarked)

512x512 pixels at 0.46um/pixel (20x)
**Situation A**

Marker is present, but underlying image is discernable by a human (enough information exists to restore image to recognize cellular objects)

![Contrast / brightness adjustment](image)

**Situation B**

Marker is present, underlying image information is mostly gone
Situation B: We don’t want GANs to make up data!

https://dribbble.com/shots/3798731-Mona-Liza-800-X600
This is better.

I’m sorry, I really didn’t know what to put here, but at least I didn’t take a wild guess?
Train the GAN to return black pixels if it does not have enough information.

Marked image  Original Image + modification
Training details

Augmentation: random cropping and rotations

GAN model: conditional GANs (require pairs of images for training) - pix2pix model.

Jun-Yan Zhu and Taesung Park
https://github.com/junyanz/pytorch-CycleGAN-and-pix2pix
Parameters: 40 epochs on 4 GPUs
Results

Unseen marked image

GAN-restored image

Original, unmarked image
Unseen marked image

GAN-restored image

Original, unmarked image
Unseen marked image

Original, unmarked image (but with the imposed black)

GAN-restored image
Unseen marked image

GAN-restored image

Original, unmarked image
Measuring GAN Error Rate and Severity

Need to measure algorithm performance quantitatively in order to assess improvement.

2 types of discrepancies between original image and GAN

- Position (where objects in the image are)
- Color balance (notably red vs blue)
  - H&E = hematoxylin/eosin
  - Information is in balance between pink and purple
Quantify error in position

Quantify difference in pixel positions: Abs [Original Image (ground truth) – GAN generated image ]
average of RGB channels = grayscale

Original

GAN

-  = 10.168
Quantification of GAN error rate: Position

These images were glass areas with marker

~18%

Position diff
GAN can remove marker from glass... Mostly...

Unseen marked image

Original, unmarked image

GAN-restored image
These images were glass areas with marker

~18%

These images had other issues with the ground truths (not registered, out of focus)

~4%
Registration problem

Blurred

Unseen marked image

Original, unmarked image

GAN-restored image

Blurred region

Unseen marked image

Original, unmarked image

GAN-restored image
These images were glass areas with marker

~18%

Most images (2.5 < x<25)
78%

Mean ~ 10

These images had other issues with the ground truths (not registered, out of focus)

~4%
Quantify error in color

Quantify different in color: Abs [Original Image (R/B) - GAN generated Image (R/B) ]
Measure change in Red: Blue color balance ratio

Original

Red

Blue

GAN

Red

Blue

\[ \text{Abs} \left( \frac{\text{Original Image}}{\text{GAN generated Image}} \right) = 0.12 \]
Quantification of GAN error rate: Red/Blue balance

Color diff
Color diff

Position diff

Original image  GAN restored image
Conclusions and Future Plans

Likely more work is needed before restored images can be used for other analyses.

- Some errors are due to ground truth errors with scanning (out of focus) and poor registration
  - Correct manually, or algorithm to specifically fix this
- Works quite well for green pen-marks, less for blue and red
- Most serious errors come from color balance.
  - Eg. Blue markers => “hematoxylin-rich” fake cells appear.
  - Maybe train with greyscale images
- Algo needs to return more black areas where it is uncertain
  - Re-train with more black areas
Conclusions and Future Plans

• More data
  • More data augmentation
  • More marked slides
• Improve error metric
  • Incorporate human evaluation
  • Incorporate ‘machine’ evaluation
    • test if segmentation algorithms etc. perform similarly on restored images compared to ‘real’ images

• Improve algorithm
  • Parameter tuning
  • Try other GAN networks
Acknowledgements

Jennifer Giltnane
Aicha BenTaieb
Quincy Wong
Priti Hegde
Craig Cummings
Andrew Whitehead
Doing now what patients need next
Previous work with GANs in biomedical space

Generation of virtual lung single-photon emission computed tomography/CT fusion images for functional avoidance radiotherapy planning using machine learning algorithms

- BumSup Jang and Ji Hyun Chang, J Medical Imaging and Radiation Oncology. 2019

SHIFT: speedy histopathological-to-immunofluorescent translation of whole slide images using conditional generative adversarial networks

- Erik A. Burlingame, Adam A. Margolin, Joe W. Gray, and Young Hwan Chang Proc SPIE Int Soc Opt Eng. 2018
  - https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6166432/