Instituto Tecnológico de Costa Rica
Escuela de Ingeniería en Computación
Maestría en Ciencias de la Computación
Curso: Deep Learning
Profesor: Dr. José Carranza-Rojas
Valor: 15%
Proyecto en parejas

Semestre 2, 2021

# Proyecto 1 – DeepPaint

Colorear imágenes ha sido una tarea típicamente difícil para métodos tradicionales de Machine Learning, Esto porque coloreo realista de imágenes conlleva entender contexto de la imagen. Sin embargo, nuevos métodos usando Deep Learning han permitido generar resultados realistas que son no sólo plausibles sino atractivos a la vista [1, 2, 3].

Uno de los retos inherentes de este problema es el hecho de que ciertos elementos de la foto podrían ser pintados de diversas formas, por ejemplo, una manzana podría ser verde o roja, pero no azul. Esto quiere decir que existe intrínsecamente un contexto que, un modelo robusto, debería tomar en cuenta, donde cierto entendimiento de la escena debe darse para producir resultados congruentes.

Adicionalmente, el problema de generar colores vívidos y no necesariamente apagados existe. Generalmente está asociado a cómo se predicen los colores de cada pixel y en muchos casos, a qué tipo de loss function se usa para entrenar al modelo y calcular su error. Algunas loss functions como el l2 norm suele generar colores muy poco vívidos, no apetecibles al avista, e incluso a veces, no plausibles.

Este proyecto se basa en construir una arquitectura, un loss function, y un módulo (capa) de Deep Learning, que permite “pintar” de forma coherente una imagen de tonalidades grises. El estudiante deberá hacer un ablation study para entender cómo cada parte de la metodología afecta el rendimiento del modelo.

## Sets de datos

PASCAL VOC dataset:

Este set de datos contiene alrededor de 11k~ fotos de entrenamiento y 7k~ fotos de testing. En general contiene alrededor de 20 clases y tiene metadatos de segmentación que podrían ser útiles para otras tareas.

http://host.robots.ox.ac.uk/pascal/VOC/voc2012/index.html

Es importante mencionar que se recomienda pasar las fotos de RGB a CIE Lab Color Space, el cual usar 3 canales distintos que RGB. Uno de ellos, el L, es justamente la foto en grises, y de esta forma el modelo debe predecir los otros dos canales, a y b. Esto hace más sencilla la tarea de colorear. En otras palabras, al convertir la imagen original

Instituto Tecnológico de Costa Rica Semestre 2, 2021
Escuela de Ingeniería en Computación
Maestría en Ciencias de la Computación
Curso: Deep Learning
Profesor: Dr. José Carranza-Rojas
Valor: 15%
Proyecto en parejas

RGB en CIE Lab, obtenemos el input del modelo (canal L), y un tensor de WxHx2 canales de label (canales a y b en un solo tensor).

## Arquitectura Base

Se espera que el estudiante monte la arquitectura basado en su nivel de conocimiento y gusto (es abierto al estudiante, en otras palabras). Lógicamente se espera ver resultados decentes, por lo que hay un mínimo de complejidad del modelo que debe cumplirse. La arquitectura base puede ir desde algo sencillo, a algo existente como UNet, hasta algo complejo como GANs. La arquitectura del modelo puede incluso tener varias redes, para los que quieren incursionar en GANs o Autoencoders. Esto queda totalmente abierto al estudiante.

Deben existir dos arquitecturas, la base que puede incluso ser ya preexistente, y la modificada por el estudiante, que contiene bloques o módulos internos diseñados (en alguna medida) por el mismo estudiante. La siguiente sección describe este módulo más a fondo.

## Módulo Propio

De forma similar a los papers de Inception [4] y, por mencionar otro, el paper de ResNet [5], en este proyecto el estudiante tendrá que diseñar un bloque/módulo especial que se use en una versión mejorada de la arquitectura base. El módulo se espera que tome en cuenta alguna lógica intrínseca de algún tema que ayude a mejorar los colores generados. Estos temas pueden ser desde teoría del color, teoría de percepción, etc.

## Loss Function (Base versus Propia)

Una forma directa de calcular el error de un modelo generativo como este, es comparar directamente pixel por pixel con un loss function tipo el l2 norm. Lamentablemente, la literatura muestra que el uso de este loss function genera imágenes de colores apagados, y en muchos casos sin sentido semántico. Este loss function servirá para correr nuestros números “base”, sin embargo, parte del proyecto es proponer un loss function que dé mejores resultados y que tome en cuenta algún otro aspecto, como la percepción humana de los objetos en la escena, o bien la plausibilidad del color de los objetos, más allá que una simple distancia entre pixels de la etiqueta y el predicho.

## Métricas

Los resultados deben ser en términos de test sets (aunque pueden mencionar los de training set). Las siguientes son las métricas base, pero pueden usar alguna otra de verlo necesario.

1. Peak Signal-to-Noise Ratio (PSNR): Es una métrica bastante usada en tareas de reconstrucción de imágenes y otras señales . Sin embargo, es una métrica que no refleja la percepción humana respecto a la imagen, por lo que hay que tener cuidado ya que un buen PSNR no necesariamente quiere decir que la imagen sea plausible o atractiva.

Instituto Tecnológico de Costa Rica Semestre 2, 2021
Escuela de Ingeniería en Computación
Maestría en Ciencias de la Computación
Curso: Deep Learning
Profesor: Dr. José Carranza-Rojas
Valor: 15%
Proyecto en parejas

2. Structural Similarity Index (SSIM): Métrica de referencia (necesita imagen predicha e imagen destino/etiqueta) que cuantifica la degradación de imágenes y toma en cuenta (un poco) la percepción humana.
3. Fréchet inception distance (FID): Es una métrica de la calidad de imágenes generadas, generalmente por GANs. Compara la distribución de las imágenes reales, con la de las imágenes generadas. En nuestro caso, las imágenes generadas pintadas, con las de color real.

## Ablation Study

En ciencia existe el concepto de Ablation Study cuando tenemos versiones incrementales de cierta metodología, y queremos saber cómo afecta cada nueva pieza de nuestra metodología los resultados. En nuestro caso, vamos a tener una arquitectura base (con solo capas básicas de convolution, pooling, etc), seguida por una arquitectura que hace uso de bloques/módulo propio diseñados por el estudiante, y seguidamente usando el loss function propio. Se deben medir las métricas en cada experimento, para ver cómo incrementalmente el módulo propio y el loss function ayudan a mejorarlas (o empeorarlas?).

## Herramientas

Se hará uso ya sea de Keras/Tensorflow o de Pytorch para crear los modelos. Se deben entregar los fuentes con los resultados en un Jupyter Notebook. Se recomienda hacer uso de matplotlib para generar los plots. El Jupyter Notebook debe traer todos los resultados sin necesidad de correrlo. Se debe exportar el notebook a HTML también.

### Paper en Latex / PDF

Deben crear en latex (y entregar los fuentes de latex) un paper donde describen los experimentos realizados, su metodología respecto a hiperparámetros, así como loss functions, módulos, arquitectura, ablation study, etc.. Deben incluir los resultados y por supuesto sus propias conclusiones. El documento debe seguir el template de la IEEE para publicaciones en ingeniería el cual pueden encontrar aquí:

https://www.ieee.org/conferences/publishing/templates.html

### Evaluación

| Tarea | Puntaje Máximo |
|---|---|
| Baseline (arquitectura base, loss function l2 norm base) | 10 |
| Diseño y programación de módulo especial | 20 |
| Diseño y programación de loss function nueva | 20 |
| Ablation Study | 30 |
| Paper:<br>● Metodología<br>● Análisis general | 20 |

Instituto Tecnológico de Costa Rica
Escuela de Ingeniería en Computación
Maestría en Ciencias de la Computación
Curso: Deep Learning
Profesor: Dr. José Carranza-Rojas
Valor: 15%
Proyecto en parejas

Semestre 2, 2021

| | |
|---|---|
| • Completitud<br>• Descripción de Resultados<br>• Conclusiones (no solo mostrar resultados, sino interpretar)<br>• Trabajo futuro | |
| TOTAL | 100 |

## Referencias


[1] PZhang R., Isola P., Efros A.A. (2016) Colorful Image Colorization. In: Leibe B., Matas J., Sebe N., Welling M. (eds) Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science, vol 9907. Springer, Cham. https://doi.org/10.1007/978-3-319-46487-9_40

[2] Kumar, M., Weissenborn, D., & Kalchbrenner, N. (2021). Colorization Transformer. *ArXiv, abs/2102.04432*.

[3] P Vitoria, L Raad, C Ballester . “Chromagan: Adversarial picture colorization with semantic class distribution”, The IEEE Winter Conference on Applications of Computer Vision, 2020

[4] C. Szegedy *et al*., "Going deeper with convolutions," *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1-9, doi: 10.1109/CVPR.2015.7298594.

[5] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.