

Real-time user-guided image colorization with learned deep priors

Conrado, Wennys e Ygor

IFGoiano

24 de Setembro de 2019

Inteligência artificial pode ser definida como sendo o desafio de criar máquinas inteligentes. Ou seja, máquinas capazes de resolver problemas que tipicamente são resolvidos por humanos.

A inteligência artificial (IA) possibilita que máquinas aprendam com experiências, se ajustem a novas entradas de dados e performem tarefas como seres humanos. A maioria dos exemplos de IA sobre os quais você ouve falar hoje – de computadores mestres em xadrez a carros autônomos – dependem de deep learning e processamento de linguagem natural. Com essas tecnologias, os computadores podem ser treinados para cumprir tarefas específicas ao processar grandes quantidades de dados e reconhecer padrões nesses dados.

Machine Learning (Aprendizado de máquina)

O aprendizado de máquina (em inglês, machine learning) é um método de análise de dados que automatiza a construção de modelos analíticos. É um ramo da inteligência artificial baseado na ideia de que sistemas podem aprender com dados, identificar padrões e tomar decisões com o mínimo de intervenção humana.

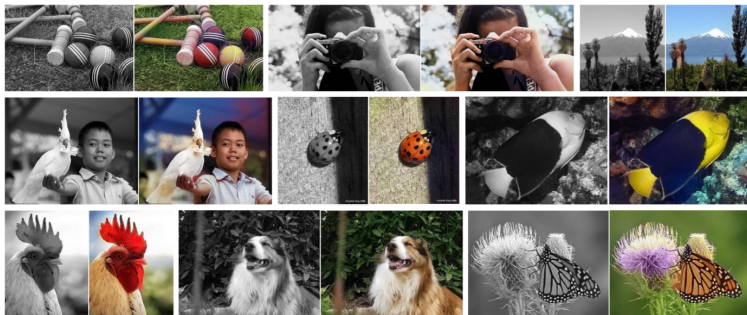
Deep Learning (Aprendizagem profunda)

Deep learning é um tipo de machine learning que treina computadores para realizar tarefas como seres humanos, o que inclui reconhecimento de fala, identificação de imagem e previsões. Em vez de organizar os dados para serem executados através de equações predefinidas, o deep learning configura parâmetros básicos sobre os dados e treina o computador para aprender sozinho através do reconhecimento padrões em várias camadas de processamento.

Há algo único e poderosamente satisfatório sobre o simples ato de adicionar cor as imagens preto e branco. Seja como uma forma de reviver memórias passadas e esquecidas ou de expressar manifestações artísticas e sua criatividade.

Dada uma fotografia em escala de cinza como entrada, este artigo ataca o problema de alucinar uma versão em cores plausível da fotografia. Propomos uma abordagem totalmente automática que produz cores vibrantes e realistas. Abraçamos a incerteza subjacente do problema, colocando-o como uma tarefa de classificação e usamos o reequilíbrio de classe no tempo do treinamento para aumentar a diversidade de cores no resultado.

Objetivo



Definindo o problema de colorização:

À primeira vista, alucinar suas cores parece assustador, pois muitas informações (duas das três dimensões) foram perdidas. Observando com mais atenção, porém, percebe-se que, em muitos casos, a semântica da cena e sua textura superficial fornecem pistas amplas para muitas regiões de cada imagem: a grama é tipicamente verde, o céu tipicamente azul e a joaninha é definitivamente vermelho. Obviamente, esse tipo de precedentes semânticos não funciona para tudo, por exemplo, as bolas de croquet na grama podem não ser, na realidade, vermelhas, amarelas e roxas (embora seja um palpite bastante bom).

Definindo problema de Colorização

No entanto, nosso objetivo não é necessariamente recuperar a cor real da verdade do solo, mas produzir uma coloração plausível que possa enganar um observador humano. Portanto, nossa tarefa se torna muito mais viável: modelar o suficiente das dependências estatísticas entre a semântica e as texturas das imagens em escala de cinza e suas versões de cores para produzir resultados visualmente atraentes.

Definindo o problema de Colorização

A previsão de cores é inerentemente multimodal - muitos objetos podem assumir várias colorizações plausíveis. Por exemplo, uma maçã é tipicamente vermelha, verde ou amarela, mas é improvável que seja azul ou laranja. Para modelar adequadamente a natureza multimodal do problema, prevemos uma distribuição de cores possíveis para cada pixel.

Definindo o problema de colorização:

Normalmente, a forma como uma imagem é codificada digitalmente é feita usando o modelo RGB. Esse modelo de cores funciona de forma aditiva, onde o vermelho (Red), verde (Green) e azul (Blue) são adicionados de várias maneiras, para reproduzir uma ampla variedade de cores.

Definindo o problema de colorização:

No entanto, o modelo usado nesse projeto se chama CIELAB. Nesse modelo, as cores são expressas em três valores numéricos.

O canal L codifica a informação referente a intensidade luminosa.

O canal A codifica os componentes verde e vermelho.

O canal B codifica os componentes azul e amarelo.

Método conhecido como Convolutional Neural Networks (CNN) ou em português Rede Neural Convolutacional, é uma classe de rede neural do tipo feed-forward, onde as imagens são convertidas em blocos menores permitindo um rápido processamento. No caso deste modelo em particular são utilizadas camadas conv, ReLU e seguidas por uma camada BatchNorm.

O termo tem origem na língua inglesa e significa “olhar para a frente” ou “avançar”. É com base nesse olhar direcionado para o horizonte que o feedforward se baseia para orientar as pessoas sobre as habilidades que precisam ser desenvolvidas. Tem como função buscar otimizar o potencial de cada profissional. E, assim, monitorar o processo de desenvolvimento para o futuro.

Como são feitas?

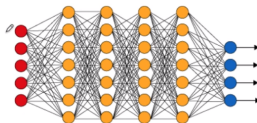
Basicamente as imagens são transformadas em valores numéricos e em seguida passam por transformações ou “convoluções” que fazem a identificação de padrões nas imagens. O algoritmo identifica os objetos na imagem e aplica as cores.

Como São Feitas

- Laranja camisa
- Azul calção
- Azul sapato



8.97, 3.45, 2.35, 0.0, 00.00, 0.00, Bart
6.75, 0.94, 0.52, 0.00, 0.00, 0.00, Bart
9.69, 4.10, 1.56, 0.00, 0.00, 0.00, Bart
0.00, 0.00, 0.00, 4.68, 0.66, 0.01, Homer
0.00, 0.00, 0.00, 0.12, 2.50, 0.03, Homer
0.00, 0.00, 0.00, 5.80, 0.50, 1.28, Homer



- Marrom boca
- Azul calca
- Cinza sapato



Como São Feitas?

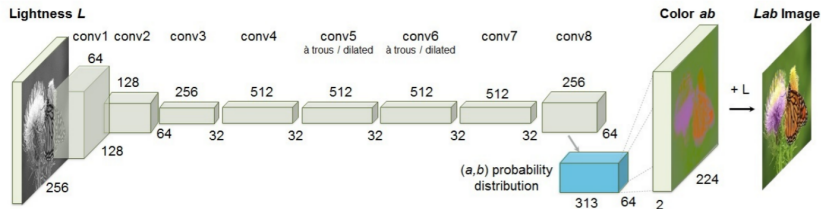


Fig. 2. Nossa arquitetura de rede. Cada camada conv refere-se a um bloco de 2 ou 3 repetidos camadas conv e ReLU, seguidas por uma camada BatchNorm [30]. A rede não possui camadas de pool. Todas as alterações na resolução são obtidas através de downsampling ou upsampling espacial entre blocos de conv.

O NumPy é o pacote fundamental para a computação científica com Python. Ele contém entre outras coisas:

um poderoso objeto de matriz N-dimensional

-Funções sofisticadas (transmissão)

-Ferramentas para integrar código C / C ++ e Fortran

-Recursos úteis de álgebra linear, transformação de Fourier e números aleatórios

Além de seus óbvios usos científicos, o NumPy também pode ser usado como um eficiente recipiente multidimensional de dados genéricos. Tipos de dados arbitrários podem ser definidos. Isso permite que o NumPy se integre de maneira fácil e rápida a uma ampla variedade de bancos de dados.

É uma biblioteca multiplataforma, totalmente livre ao uso acadêmico e comercial, para o desenvolvimento de aplicativos na área de Visão computacional

O OpenCV possui módulos de Processamento de Imagens e Video I/O, Estrutura de dados, Álgebra Linear, GUI (Interface Gráfica do Usuário) Básica com sistema de janelas independentes, Controle de mouse e teclado, além de mais de 350 algoritmos de Visão computacional como: Filtros de imagem, calibração de câmera, reconhecimento de objetos, análise estrutural e outros. O seu processamento é em tempo real de imagens.

Argparse é uma biblioteca completa de processamento de argumentos. As ações suportadas incluem armazenar o argumento (individualmente ou como parte de uma lista), armazenar um valor constante quando o argumento é encontrado (incluindo tratamento especial para valores verdadeiros / falsos para comutadores booleanos), contando o número de vezes que um argumento é visto, e chamando um retorno de chamada.