

# Pintura Artística

Dalai S. Ribeiro<sup>1</sup>, Gleidson Mendes<sup>1</sup>,

<sup>1</sup>Universidade Federal do Maranhão (UFMA)  
Av. dos Portugueses, 1966 - Bacanga, São Luís - MA, 65080-805

{dalai.ribeiro, gleidson.mendes}@gmail.com

**Abstract.** *In this article we introduce an artificial system based on a Convolutional Neural Networks that creates artistic images of high perceptual quality. The system uses neural representations to separate and recombine content and style of arbitrary images, providing a neural algorithm for the creation of artistic images.*

**Resumo.** *Neste artigo apresentamos um sistema artificial com base em Redes Neurais Convolucionais que cria imagens artísticas de alta qualidade perceptual. O sistema utiliza representações neurais para separar e recombinar conteúdo e estilo de imagens arbitrárias, proporcionando um algoritmo neural para a criação de imagens artísticas.*

## 1. Introdução

Na artes plásticas, especialmente a pintura, os seres humanos têm dominado a habilidade para criar experiências visuais únicas através da composição de uma complexa interação entre o conteúdo eo estilo de uma imagem. Até agora a base algorítmica do presente processo é desconhecido e não existe nenhum sistema artificial com capacidades semelhantes. []

## 2. Objetivos

O presente artigo visa reproduzir o trabalho [nome do trabalho], de [citar] e analisar seus resultados como forma de exercitar os conhecimentos adquiridos da disciplina de Visão Computacional do curso de Ciência da Computação da Universidade Federal do Maranhão. O artigo de [citar] introduz um sistema artificial que imagens artísticas de alta qualidade percebida. O sistema usa representações neurais para separar e recombinar conteúdo e estilo de imagens arbitrárias, criando um algoritmo neural para a criação de imagens artísticas.

## 3. Referencial Teórico

### 3.1. Aprendizado de Máquina e Redes Neurais

Aprendizado de Máquina é uma área da Inteligência Artificial cujo objetivo é o desenvolvimento de técnicas computacionais sobre o aprendizado bem como a construção de sistemas capazes de adquirir conhecimento de forma automática. Um sistema de aprendizado é um programa de computador que toma decisões baseado em experiências acumuladas através da solução bem sucedida de problemas anteriores. [Gatys et al. 2015]

Entre as diversas variações de Aprendizado de Máquina, uma das qual se destacam é as Redes Neurais. Redes Neurais, podem ser definidas intuitivamente como modelos matemáticos inspirados no funcionamento dos neurônios biológicos e na estrutura do cérebro.

### **3.2. Redes Neurais Convolutivas**

A classe de Redes Neurais que são mais poderosos em tarefas de processamento de imagem são chamados de Redes Neurais Convolutivas (RNC). RNC's consistem em camadas de pequenas unidades computacionais que processam informação visual hierarquicamente em feed-forward. Cada camada de unidades pode ser entendido como um conjunto de filtros de imagem, cada um dos quais extrai uma certa característica da imagem de entrada. Assim, a saída de uma determinada camada consiste dos chamados mapas de características: diferentes versões filtradas da imagem de entrada.

Quando Convolucionais Redes Neurais são treinados sobre reconhecimento de objetos, eles desenvolvem uma representação da imagem que faz a informação do objeto cada vez mais explícita ao longo da hierarquia.<sup>8</sup> Portanto, ao longo da hierarquia de processamento da rede, a imagem de entrada é transformado em representações que cada vez mais se preocupam com o conteúdo real da imagem em relação aos seus valores de pixel detalhados

## **4. Método**

Os resultados apresentados no artigo foram gerados em função da VGG-Network, uma Rede Neural Convolucional que rivaliza com o desempenho humano na tarefa de reconhecer um objeto visual comum e foi introduzido e amplamente descrito no artigo [nome do artigo] por [Simonyan 2014]. O autor utiliza o mapa de características oferecida pelas 16 camadas convolutivas 5 camadas de pooling da VGG-Network. Não foi utilizado nenhuma das camadas completamente conectadas.

As camadas mais altas da RNC capturam o conteúdo de alto nível em termos de objetos e sua disposição na imagem de entrada, sem se limitar aos valores de pixel exatos da reconstrução. Em contraste, as reconstruções das camadas inferiores reproduzem simplesmente os valores de pixel exatos da imagem original. Nós, portanto, referimos as imagens geradas nas camadas mais altas da rede como a representação do conteúdo.

Para obter uma representação do estilo de uma imagem de entrada, usamos um mapa de características originalmente projetado para capturar informação de textura. Este mapa é construído em cima das respostas dos filtros em cada camada da rede. Consiste nas correlações entre as diferentes respostas do filtro sobre a extensão espacial dos mapas de características. Ao inclui-las, obtém-se uma representação estacionária, multi-escala de imagem de entrada, que capta a sua informação de textura, mas não o arranjo global.

A principal contribuição deste trabalho é que as representações de conteúdo e estilo na Rede Neural Convolucional são separáveis. Ou seja, nós podemos manipular ambas as representações de forma independente para produzir novas imagens, perceptivelmente significativas.

As imagens são sintetizadas achando uma imagem que corresponde ao mesmo tempo o teor de representação da fotografia e o estilo de representação da respectiva peça de arte. Embora o arranjo global da fotografia original é preservada, as cores e as estruturas locais que compõem o cenário mundial são fornecidos pela obra de arte.

Efetivamente, isso torna a fotografia no estilo da obra de arte, de modo que a aparência da imagem sintetizada assemelha-se a obra de arte, mesmo que ele mostra o mesmo conteúdo da fotografia.

Para a síntese de imagem o autor substituiu a operação max-pooling por pooling média para melhorar o gradiente flow e obter resultados ligeiramente mais atraentes.

Geralmente cada camada na rede define um banco de filtros não-linear, cuja complexidade aumenta com a posição da camada na rede. Assim, uma determinada entrada imagem  $x$  é codificado em cada camada da RNC pelas respostas de filtro para essa imagem. Uma camada com filtros distintos  $N^l$  tem a característica  $N^l$  Mapas de cada um de tamanho  $M^l$ , onde  $M^l$  é a altura vezes a largura do mapa de características. Assim, as respostas em uma camada de  $L$  pode ser armazenado numa matriz  $F = R \times M^l$  Onde  $F_{ij}^l$  é a activação do filtro  $I$  na posição  $j$  na camada  $L$ . Então,  $p$  e  $x$  ser a imagem original e a imagem que é gerada e  $P$  e  $F$  a respectiva representação no mapa de caracteritscas. O autor, então, define a perda de erro quadratico entre a representação duas características representadas pela Equação 1 de [Gatys et al. 2015]

O derivado desta perda no que diz respeito às activações na camada  $l$  é igual a equação 2 de [Gatys et al. 2015]

No topo das respostas da RNC em cada camada da rede, o autor constrói uma representação estilo que computa as correlações entre as diferentes respostas de filtro, onde a expectativa é tomada sobre o espacial estender da imagem de entrada. Estas correlações característica são dadas pela matriz Gram  $G^l = R(N \times N^l)$ , onde  $G_{ij}^l$  é o produto interno entre o mapa característica vectorised  $i$  e  $j$  na camada  $l$  como utilizado na equação 3 de [Gatys et al. 2015]

Para gerar as imagens que misturam o conteúdo de uma fotografia com o estilo de uma pintura minimizamos conjuntamente a distância de uma imagem de ruído branco a partir da representação conteúdo da fotografia em uma camada da rede e a representação do estilo da pintura em um número de camadas do RNC. Então  $p$  é a fotografia e  $a$  é a obra de arte. A função de perda minimizada é:

## 5. Resultado

Apesar de seguir os passos apresentados pelo artigo, não foi possível apresentar resultados satisfatórios. Utilizamos a framework Caffe [referencia ao site do caffe] para o suporte na criação da rede neural convolutiva.

Para níveis de teste ao invés de utilizar o proposto com 16 camadas e average pooling, foi utilizado 3 camadas 1.

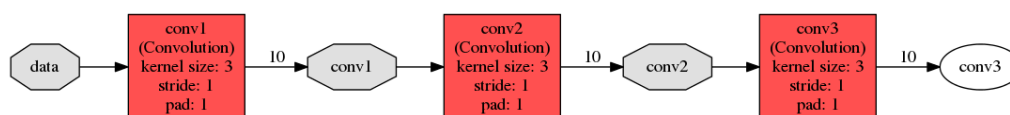


Figure 1. Rede Neural Modelo

É aplicado 2 métodos para obter uma melhor combinação das 2 imagens, forward e backward propagation na rede convolutiva. Para o método do Forward-propagation foi-se obtido as respostas desejadas. Uma imagem ao entrar na rede neural e ao passar pelas camadas é disposta em diferentes tipos de filtros. Porém o backward propagation não funcionou como desejado.



**Figure 2. Conteúdo**

Para primeiro passo é escolhido 2 imagens: Uma para se manter o conteúdo, e outra para se manter o estilo, apresentados nas figuras 2 e 3.



**Figure 3. Estilo**

De ambas é retirado o mapa de características do estilo e do conteúdo. Esse mapa de características vai ser o mapa de características padrão da imagem original.

Para cada mapa de característica padrão é feito um método de forward propagation da imagem escolhida (conteúdo ou estilo por vez). Seguindo a equação (1) e (2) do artigo de [Gatys et al. 2015] uma matriz  $M$  de tamanho número de filtros  $x$  (largura \* altura do resultado de um filtro). Assim, se em uma camada tiver 10 filtros e que a imagem resultante seja  $20x20$  teremos uma mapa de característica de  $M(10x400)$ . A imagem  $\zeta_{union}$  mostra a imagem gerada de um filtro da imagem de conteúdo e a derivada reconstruída como imagem proposta no artigo

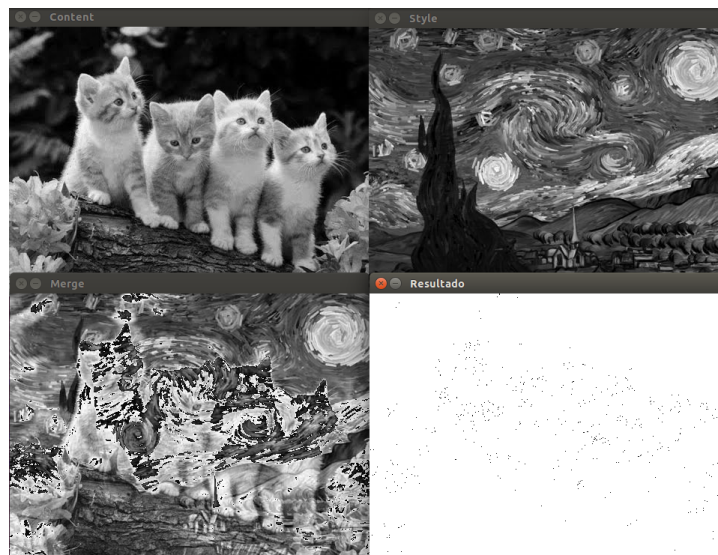


**Figure 4. União**

Para o estilo é criado para cada camada uma matriz  $M$  (número de filtros x número de filtros) e a relação entre os filtros se dá por produto interno, tratando os vetores de filtros do mapa de característica como vetores. Como apresentado nas equações (3, 4 e 5) do artigo *ARTIGO PRINCIPAL*.

Os valores obtidos (perda e gradiente) das equações devem ser retornados por backward propagation com o intuito de minimizar a diferença entre as 2 imagens (conteúdo e estilo). Assim, ao se fazer o processo de backward propagation, a imagem que iniciou a rede convolutiva agora é alterada por uma nova solução, e esta nova solução irá passar por uma nova interação de forward e backward propagation afim de se obter uma adaptação das 2 imagens.

A última equação utilizada pelo artigo é a equação para unir as 2 imagens de acordo com um peso para estilo e para conteúdo, afim de reproduzir uma nova imagem.



**Figure 5. Resultado - Janela 1:**

Em nosso trabalho, não conseguimos utilizar método backward propagation como o esperado, na adaptação dos erros. Outro problema encontrado foi no entendimento do cálculo para imagens depois de executado o average pooling uma vez que as matriz

teriam dimensões diferentes impossibilitando um cálculo entre matrizes, razão pela qual não utilizamos o modelo utilizado no artigo de estudo

## **6. References**

### **References**

Gatys, L. A., Ecker, A. S., and Bethge, M. (2015). A neural algorithm of artistic style.

Simonyan, K. . Z. (2014). A. very deep convolutional networks for large-scale image recognition.