

**FUNDAÇÃO GETULIO VARGAS
ESCOLA DE MATEMÁTICA APLICADA**

DANILO LEMOS CARDOSO

SÍNTESE DE TEXTURAS BASEADO EM EXEMPLO: SUBTITULO

Rio de Janeiro
2022

DANILO LEMOS CARDOSO

SÍNTESE DE TEXTURAS BASEADO EM EXEMPLO

Trabalho de conclusão de curso apresentada para a Escola de Matemática Aplicada (FGV/EMAp) como requisito para o grau de bacharel em Matemática Aplicada.

Área de estudo: área de estudo.

Orientador: Asla Medeiros e Sá

Rio de Janeiro

2022

Ficha catalográfica elaborada pela BMHS/FGV

Sobrenome, Nome

Síntese de Texturas Baseado em Exemplo: subtitulo/ Danilo Lemos Cardoso. –
2022.

20f.

Trabalho de Conclusão de Curso – Escola de Matemática Aplicada.

Advisor: Asla Medeiros e Sá.

Includes bibliography.

1. Matemática 2. Aplicada 2. na matemática I. Sobrenome professor, Nome professor II. Escola de Matemática Aplicada III. Síntese de Texturas Baseado em Exemplo

DANILO LEMOS CARDOSO

SÍNTESE DE TEXTURAS BASEADO EM EXEMPLO: SUBTITULO

Trabalho de conclusão de curso apresentada para a Escola de Matemática Aplicada (FGV/EMAp) como requisito para o grau de bacharel em Matemática Aplicada.

Área de estudo: área de estudo.

E aprovado em 21/12/2021
Pela comissão organizadora

Asla Medeiros e Sá
Escola de Matemática Aplicada

Convidado 1
Instituição 1

Convidado 2
Instituição 2

Resumo

Temas relacionados à análise e síntese de imagens vêm tendo um crescente aumento de interesse nos últimos anos, tanto por parte de pesquisadores, quanto pelo público geral. Isso se deve ao aumento do poder computacional e da geração de dados, que possibilitam o estudo e desenvolvimento de modelos mais ricos e com melhores resultados.

Neste trabalho será tratada a geração de texturas de tamanho arbitrário usando uma amostra limitada, tentando assim modelar seu processo de geração para produzir um resultado perceptualmente semelhante ao original. Será feita uma revisão da literatura sobre o tema, mostrando os principais resultados e abordagens, como a área foi se desenvolvendo até chegar no conhecimento de hoje, e como esse área influenciou em outros temas relacionados a imagem.

No final será mostrada uma implementação do método computacional de geração de texturas, aplicando-o em diferentes tipos de textura para observar como características da imagem original podem mudar a qualidade do resultado. Em seguida serão exploradas variações do método que permitem um melhor controle da forma do resultado final.

Palavras-chave: síntese de textura.

Sumário

1	INTRODUÇÃO	6
1.1	O que é textura	6
1.2	O que é síntese de textura	7
2	MODELOS	8
2.1	Modelos paramétricos	8
2.2	Modelos não paramétricos	10
2.3	Modelos de aprendizado profundo	10
3	DEEP LEARNING	14
3.1	Aprendizado de Representação	14
3.2	Redes neurais	14
3.3	Redes convolucionais	16
3.4	Síntese de textura com Redes Convolucionais	16
4	RESULTADOS	17
5	CONCLUSÃO	20

1 Introdução

...

O trabalho foi baseado em dois Surveys na área. O primeiro de Wei, Lefebvre, Kwatra e Turk (**Wei2009**) dá ênfase em abordagens não paramétricas e texturas dinâmicas. Já o segundo de Raad, Davy, Desolneux e Morel (**Raad2018**) mostra a síntese com métodos mais recentes usando Deep Learning e Redes Convolucionais. Analisando os dois Surveys, fica bem claro o quanto a área se modificou depois da revolução do Deep Learning, que se iniciou em 2015 com o trabalho de Yann LeCun, Yoshua Bengio e Geoffrey Hinton (**LeCun2015**).

1.1 O que é textura

A palavra “textura” pode ter diferentes significados dependendo do contexto. Um deles se refere às diferentes características da superfície de um objeto, sejam elas visuais (cor, desenhos), geométricas (relevo, forma) ou táteis (maciez, dureza). Em computação gráfica, textura geralmente é o nome que se dá a uma imagem (matriz de pixels) que descreve alguma característica da superfície de um objeto, como a cor ou a direção do vetor normal (utilizada para iluminação).

Para o processo de síntese, é preciso restringir o conjunto de imagens consideradas texturas àquelas que apresentam algum tipo de padrão perceptual em seu domínio. Com isso é possível fazer a síntese estudando e imitando o processo que gerou esse padrão.

Cross, G.R. e Jain, A.K. (**Cross1983**) descrevem textura como sendo um Campo Markiviano Aleatório (Markov Random Field, MRF). Esse modelo é o mais usado no processo de síntese pois satisfaz a propriedade de Markov (o valor de cada pixel dado sua vizinhança não depende do resto da textura) e a homogeneidade (a distribuição é invariante por translação), logo se encaixa com as necessidades descritas anteriormente.

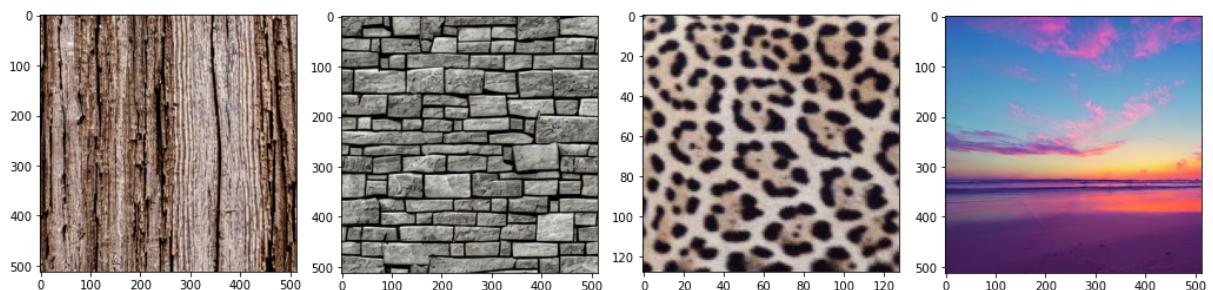


Figura 1 – As três primeiras imagens são texturas por apresentarem um padrão visual estacionário. A quarta imagem apresenta uma estrutura não estacionária, não permitindo uma extração infinita.

1.2 O que é síntese de textura

O processo de síntese de textura baseado em amostra não tem uma definição clara matematicamente, é algo mais intrínseco à percepção humana. O objetivo é, a partir de uma amostra de textura, gerar outras texturas de tamanho arbitrário que imitam o processo gerador da amostra. Esse processo gerador é baseado em métricas perceptuais, que não podem ser definidas de forma fechada, pois podem depender de aspectos finos da imagem, como forma e iluminação. Assim, os trabalhos na área nos últimos anos consistem em tentar descobrir melhores aproximações para essa métrica perceptual.

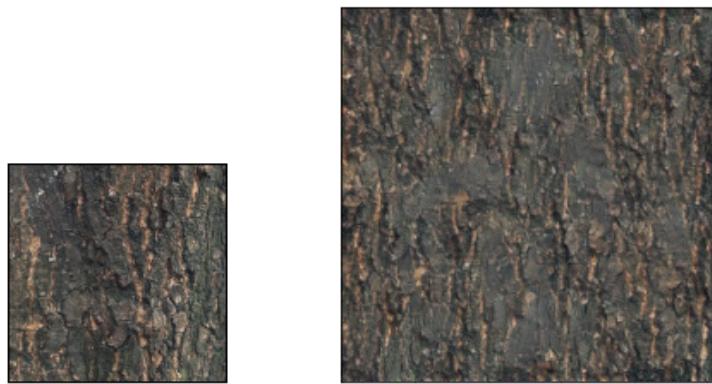


Figura 2 – Exemplo de síntese de textura. A imagem maior é visualmente semelhante a imagem menor.

Ao restringir o conjunto de imagens aos Campos Markovianos, o processo de síntese pode ser descrito como uma re-amostragem da distribuição condicional da amostra. Com isso, o desafio do método passa a ser descobrir a distribuição a partir da amostra.

2 Modelos

Com o crescimento da área de processamento de imagens e do poder computacional, os modelos de Síntese de Textura foram se aprimorando ao longo do tempo. As melhorias podem ser de qualidade do resultado ou no tempo de execução do método. Neste capítulo será apresentado um breve resumo de alguns dos principais modelos da área.

2.1 Modelos paramétricos

Os modelos que fazem a síntese a partir de um conjunto de estatísticas da amostra original são chamados modelos paramétricos. Esses modelos partem de um ruído e fazem a síntese reduzindo a diferença entre as estatísticas desse ruído e da amostra utilizando uma função de otimização. A qualidade do modelo vai depender do conjunto de estatísticas escolhido e do tipo de otimização utilizado.

Heegen e Bergen (**Heeger1995**) foram um dos pioneiros no ramo. O método parte de um ruído branco e faz a síntese aproximando os histogramas das funções de distribuição acumulada. O método usa transformações por pirâmides, que decompõe a imagem em representações em diferentes escalas, e faz a aproximação do histograma em cada uma dessas camadas.



Figura 3 – Síntese de Textura de Heegen e Bergen. O modelo conseguia sintetizar elementos estocásticos da textura.

De Bonet (**Bonet1997**) propôs um método que calcula a distribuição conjunta da amostra em diferentes escalas. Em seguida a amostragem é feita partindo de escalas mais grosseiras até escalas mais finas.

Zhu, Wu e Mumford (**Zhu1998**) utiliza um banco de filtros para obter uma representação da imagem e extrair o histograma. Em seguida deriva-se a distribuição desses filtros para serem re-amostrados usando a técnica de Gibbs Sampling.



Figura 4 – Modelo de De Bonet. O trabalho era feito escala por escala. Isso permitia sintetizar elementos mais regulares da textura.

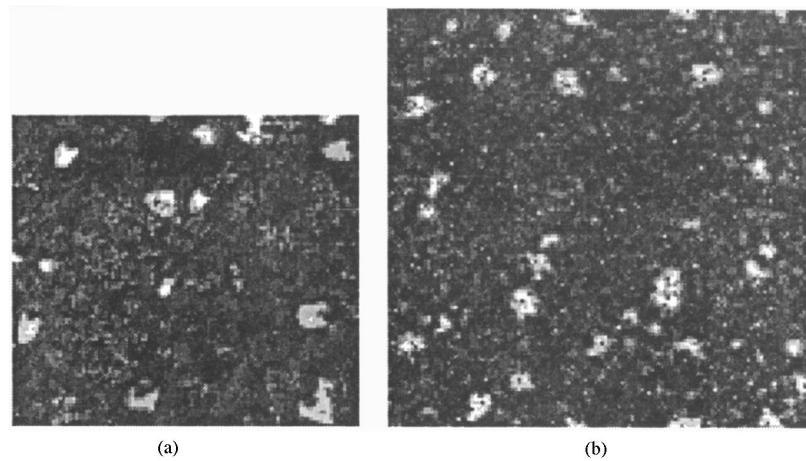


Figura 5 – Modelo de Zhu, Wu e Mumford usando 5 filtros.

Portilla e Simoncelli (**Portilla1999**) propuseram fazer a amostragem usando as autocorrelações da Transformada de Wavelet da amostra.

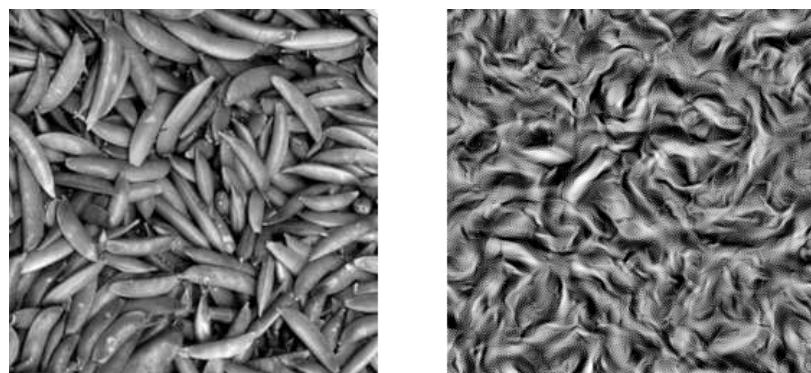


Figura 6 – Teste do modelo de Portilla e Simoncelli com uma textura mais estruturado. É possível notar que há uma dificuldade em representar tais estruturas.

Béla Julesz foi um neurocientista que trabalhava na percepção visual humana. Em

uma de suas publicações (**Julesz1981**), Julesz fez experimentos que mostravam que um conjunto de imagens com mesma informação visual podem ser representadas pela densidade do que ele chama de “Textons”, que são estruturas como cor e arestas presentes na imagem. Esse tipo de representação, embora tenha guiado o trabalho de modelos paramétricos, não deixa claro como fazer essa representação em Textons. Assim, na tentativa de encontrar tal representação, os métodos acabam sendo difíceis de implementar, e nem sempre conseguem ter sucesso em diferentes texturas.

2.2 Modelos não paramétricos

Diferentemente dos modelos paramétricos, os modelos não paramétricos não dependem do cálculo de alguma estatística da amostra original para o processo de síntese. Ele gera a imagem pegando informação diretamente da amostra de modo a simular a amostragem da textura.

Essa forma de amostragem foi proposta inicialmente por Efros e Leung (**Efros1999**). Ela consistia em re-amostrar a textura pixel por pixel, pegando diretamente da imagem original o pixel que tem a vizinhança mais parecida com a vizinhança de seu destino. Os resultados na época foram bem superiores aos que se podiam obter com os métodos paramétricos, mas a procura por todas as vizinhanças na amostra tornava o método lento.

Mais tarde, Efros e Freeman (**Efros2001**) propuseram o método de Quilting (costura), que fazia a amostragem a partir de pedaços da imagem original. O método consistia em selecionar janelas de tamanho fixo da amostra e distribuí-las com uma sobreposição sobre elas. Em seguida é usado um algoritmo de min-cut para dividir essas janelas em pedaços que se encaixam para formar a nova textura. Essa abordagem era mais rápida computacionalmente do que o método anterior, e produzia resultados tão bons quanto.

Com os avanços na área, Vivek Kwatra (**Kwatra2005**) propôs uma amostragem fazendo a minimização do que ele define como função de energia. Essa função é a diferença quadrática entre as vizinhanças da textura gerada e as vizinhanças mais próximas de cada uma. O método usa uma variação do algoritmo EM, onde na faze “E” a energia é diminuída por mínimos quadrados, e na faze “M” a energia é diminuída escolhendo as vizinhanças mais próximas na amostra.

2.3 Modelos de aprendizado profundo

Os modelos de aprendizado profundo se aproveitam dos avanços recentes que a área de Aprendizagem de Máquina para melhorar o processo de síntese. Os métodos a seguir utilizam as representações geradas por redes treinadas para a detecção de objetos

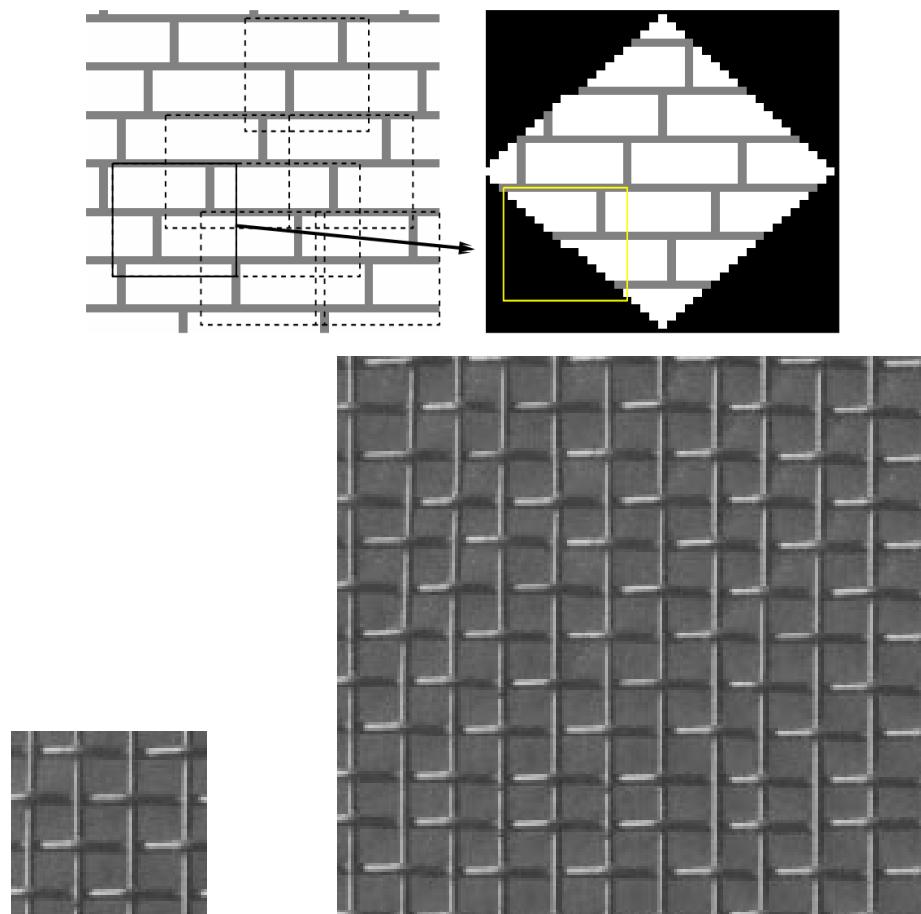


Figura 7 – O algorítmo de Efros e Leung vai escolher o pixel com vizinhança mais parecida na amostra.

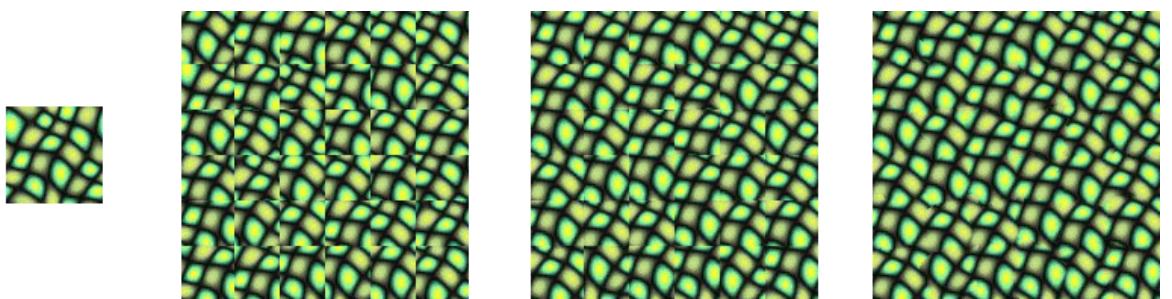


Figura 8 – O modelo Quilting de Efros e Freeman escolhe pedaços aleatórios da amostra e encaixa os quadrados.

para melhorar métodos paramétricos de síntese já existentes. As representações geradas por essas redes conseguem sintetizar bem a informação visual da imagem, oferecendo uma boa métrica perceptual.

O primeiro a propor Síntese de Textura baseado nesses modelos foram Gatys, Ecker e Bethge (**Gatys2015**). O trabalho utiliza as autocorrelações da textura semelhante ao que foi proposto por Portilla e Simoncelli, porém, em vez de utilizar transformações da amostra escolhidas a mão, utilizam a saída da rede convolucional pré treinada. Esse método será tratado com detalhes nesse trabalho, seguindo de uma implementação.

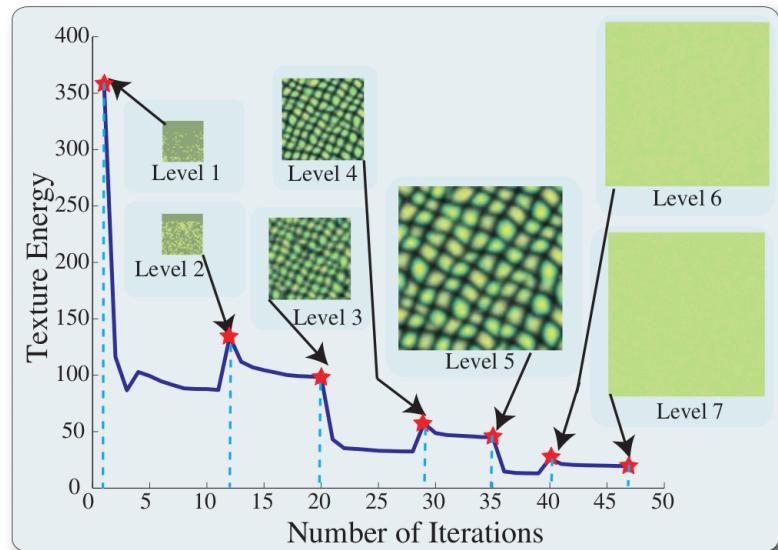


Figura 9 – Modelo de Vivek Kwatra. A função de energia cai a cada nível de resolução.



Figura 10 – Modelo de Gatys, Ecker e Bethge que utiliza as matrizes de Gram das convoluções. O método permite não ficar muito preso na amostra e produz resultados mais criativos.

Lu, Zhu e Wu (**Lu2016**) utilizam as representações das redes convolucionais no lugar do banco de filtros no método de Zhu, Wu e Mumford, produzindo melhores resultados sem se preocupar com o tipo de filtros escolhidos.

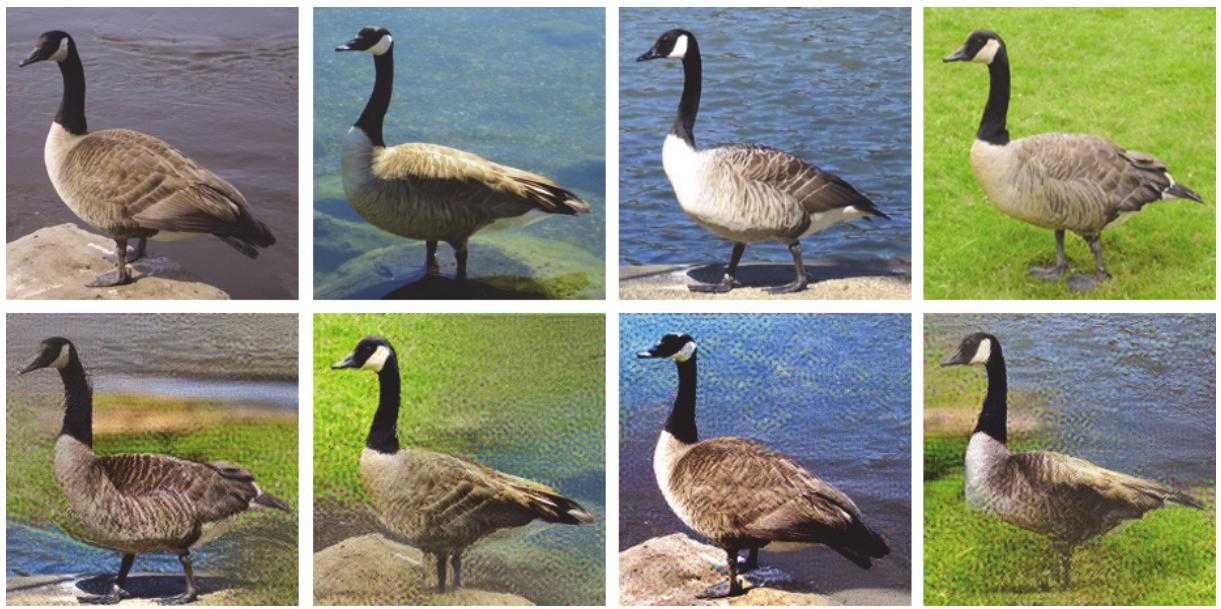


Figura 11 – Modelo de Lu, Zhu e Wu que pega quatro imagens como entrada (as de cima) e gera quatro imagens de saída (as de baixo).

3 Deep Learning

Um dos clássicos problemas na área de classificação de imagens era o reconhecimento de dígitos escritos à mão. A dificuldade nesse tipo de problema vinha pela grande variação de forma e posição dos números, tornando os modelos muito complexos e pouco precisos. O desenvolvimento recente na área de Machine Learning permitiu a criação de modelos simples capazes de aprender com os dados e fazer generalizações precisas, em alguns casos superando a percepção humana.

Neste capítulo serão apresentadas técnicas que usam Redes Neurais, como elas foram desenvolvendo ao longo dos anos, e como podemos aplicá-las na Síntese de Textura.

3.1 Aprendizado de Representação

No processo de análise de imagens, precisamos criar representações que melhor explicam seu conteúdo. Essas representações podem ser organizadas hierarquicamente, onde a primeira identifica arestas, a segunda indica cantos, etc. Ao descer no nível das representações queremos ter uma melhor informação semântica da imagem.

Para um problema de classificação de imagens, é preciso saber quais informações são úteis para obter uma melhor representação de seu conteúdo. Uma representação boa esconde informações redundantes e realçar fatores que melhor explicam a imagem. Bengio, Courville e Vincent (**Bengio2014**) falam do importante papel que a aprendizagem de representações tem em Machine Learning. Um algoritmo que aprendesse tais representações eliminaria o trabalho de desenvolvê-las, o que tornaria mais rápido a criação de aplicações.

3.2 Redes neurais

Redes Neurais são um tipo de modelo de Machine Learning inspirado em redes de neurônios naturais e como eles se comunicam. Elas podem ser representadas por um grafo, onde cada nó representa um valor e cada aresta representa uma dependência. Essas redes podem ter vários formatos dependendo do tipo de aplicação.

Um exemplo comum de Rede Neural é o Multi Layer Perceptron. Nesse modelo, a rede é dividida em camadas de diferentes tamanhos, e o dado flui da camada de entrada até a saída, que pode representar uma classificação da entrada ou uma função geral.

Cada camada da rede tem como parâmetros os offsets \mathbf{b}_m e uma matriz de pesos $\mathbf{W}_{m \times n}$. Assim, a transformação da entrada \mathbf{x}_n na saída \mathbf{y}_m pode ser representada com a

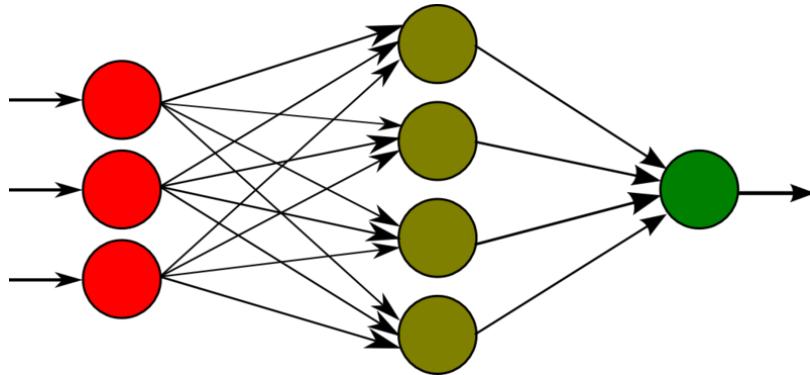


Figura 12 – Rede Multi Layer Perceptron. Cada camada da rede atua como uma representação do dado de entrada, que é usada para calcular a próxima camada.

seguinte operação:

$$\mathbf{y} = \varphi(\mathbf{W}\mathbf{x} + \mathbf{b}), \quad (3.1)$$

onde φ é uma função não linear, comumente chamada de “função de ativação”. A função não linear mais comumente usada é a ReLU, que trunca valores negativos no 0.

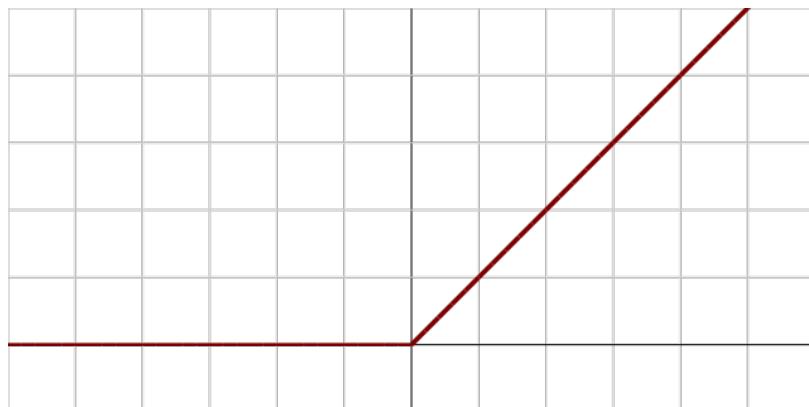


Figura 13 – Função ReLU. O seu bico não diferenciável facilita a distorcer o espaço de entrada.

Os parâmetros são calculados minimizando uma função de erro na saída. Essa minimização é feita usando o gradiente dos parâmetros em relação à função de perda. Como a derivada de todas as funções aplicadas é conhecida, o gradiente pode ser calculado computacionalmente utilizando “backpropagation”.

Esse modelo é útil por envolver operações bem simples e rápidas de serem calculadas computacionalmente. Cada camada distorce o espaço de entrada para torná-lo linearmente separável no final. Pela sua flexibilidade, o modelo geralmente é chamado de approximador universal de qualquer função, pois é capacidade de aprender funções complexas a partir de amostras.

...

3.3 Redes convolucionais

O Multi Layer Perceptron, quando aplicado diretamente nos pixels de uma imagem aparecem dois problemas consideráveis: a quantidade de parâmetros cresce muito, e informações espaciais da imagem são perdidas. Geralmente um objeto que queremos identificar pode aparecer em diferentes posições da imagem, e ao transformá-la em um vetor, perdemos essa invariância por translação. Com base nisso se tornou necessário desenvolver uma rede que aproveite a informação espacial da imagem e que tenha poucos parâmetros.

...

3.4 Síntese de textura com Redes Convolucionais

...

4 Resultados

Foi feita a implementação do método de Síntese de Textura usando a abordagem das matrizes de Gram. Para isso foi utilizada a linguagem Python com a biblioteca PyTorch. O PyTorch reúne uma série de funções que facilitaram a implementação do método. As principais foram: implementação nativa da rede VGG-19, operações aritméticas aceleradas na GPU, cálculo automático de gradiente, e implementação de métodos de otimização.

A primeira ideia era escolher um conjunto de textura de resolução 128×128 e tentar gerar imagens de 256×256 pixels. Foram escolhidos um conjunto de diferentes tipos de texturas para explorar o comportamento do algoritmo.

Em todos os testes a primeira imagem é a amostra, e as seguintes são passos (não necessariamente consecutivos) do algoritmo de otimização.

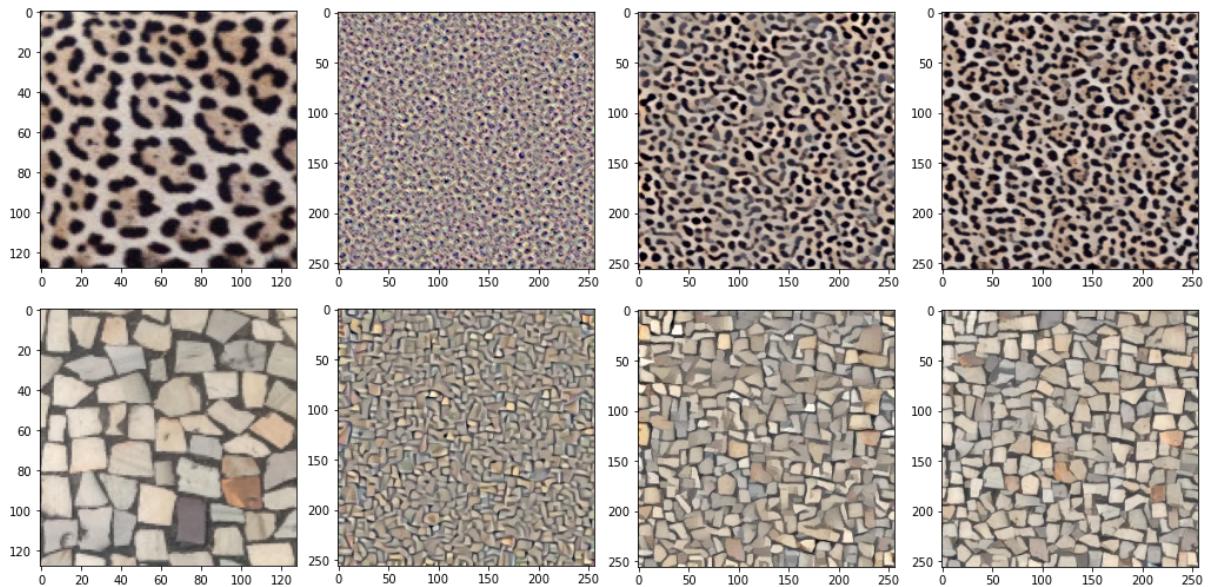


Figura 14 – Teste com texturas mais caóticas. No processo, a escala dos objetos se mantém, portanto o resultado terá quatro vezes mais informação.

Ao gerar algumas texturas, notou-se que durante o processo de otimização a imagem fica preza em uma estrutura específica. Com isso foi pensado em adicionar ruído na imagem de vez em quando para forçar a mudança do ponto de convergência. Isso gerou um resultado interessante de movimentar a textura a cada aplicação do ruído.

Não é preciso se limitar a texturas com esse tipo de implementação. É possível verificar o que acontece no resultado com imagens não textura.

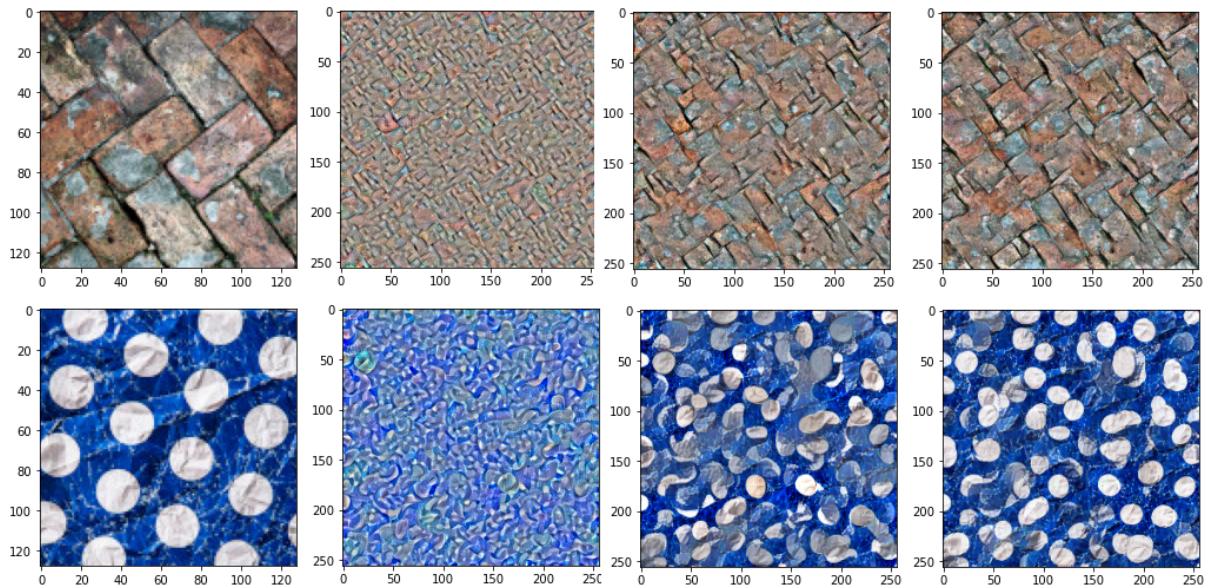


Figura 15 – O método não funciona bem em textura com estrutura regular global, ele dá preferência por estruturas locais.

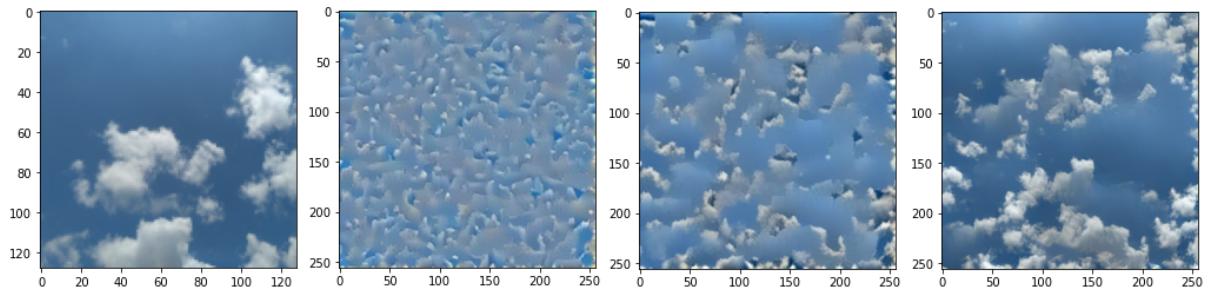


Figura 16 – Nuvens apresentam uma estrutura caótica ideal para esse método.

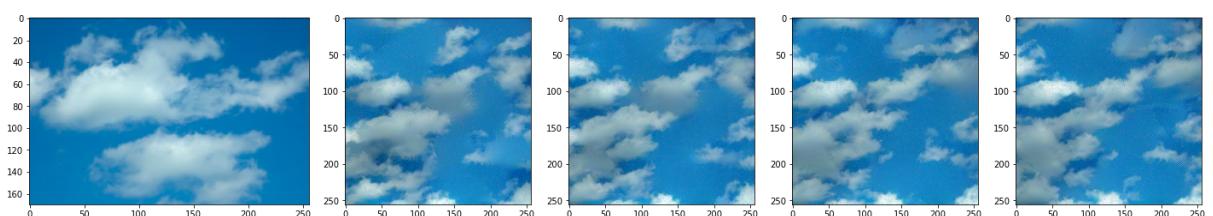


Figura 17 – A terceira imagem foi gerada a partir da adição de ruído na segunda, assim como da terceira pra quarta e da quarta pra quinta. Com isso, as quatro imagens geradas apresentam o processo de movimento das nuvens.

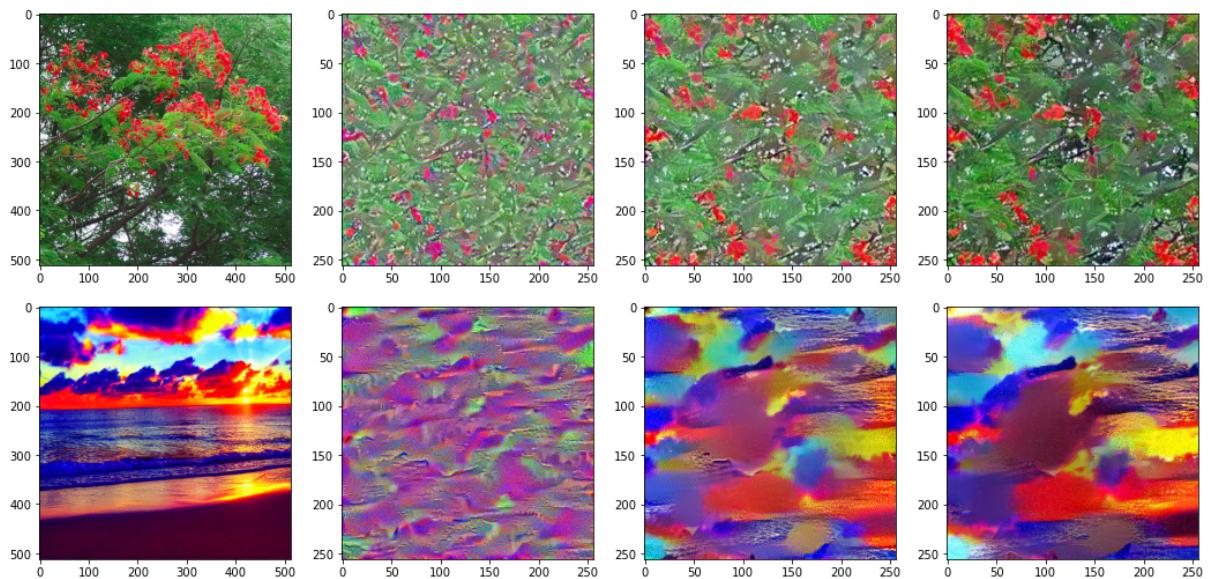


Figura 18 – A informação espacial da imagem foi perdida, mas cores e formas prevalecem.

5 Conclusão

Esse trabalho mostrou o quanto pode ser difícil a tarefa+ de processamento e síntese de texturas e imagens no geral, e o quanto foi preciso andar para chegar nos avanços que existem hoje. Uma grande quantidade de trabalhos são publicados todos os anos na área, cada um tentando encontrar uma maneira nova de melhorar a solução do problema.

Redes Neurais e Deep Learning vêm se mostrando ferramentas bem poderosas para o trabalho com imagem. Redes Convolucionais pré-treinadas para a detecção de objetos oferecem uma excelente métrica perceptual, abrindo caminho para diversos novos trabalhos. O aprendizado automático de representações facilita o trabalho de criar aplicações de processamento e síntese de imagens, mas a disponibilidade de dados e de processamento ainda pode ser um problema.

Ferramentas para trabalhar com Machine Learning como o PyTorch vêm se mostrando mais poderosas a cada dia, oferecendo facilidades para trabalhar com grandes quantidades de dados, além de fácil acesso à GPU, que melhora bastante a velocidade de operações aritméticas. Essas ferramentas também contam com sistemas de cálculo automático de gradiente, o que torna o trabalho de otimização mais fácil e menos suscetível a erros.

"