

Aula 10

Transferência de Estilo



Eduardo L. L. Cabral



Objetivos

- Definição de transferência de estilo.
- O que as RNAs convolucionais aprendem.
- Método para transferir estilo entre imagens usando RNAs.
- Função de custo utilizada na transferência de estilo.

O que é transferência de estilo

- Transferência de estilo é uma das aplicações mais interessantes de redes neurais.
- Referência \Rightarrow Gatys et al., A neural algorithm of artistic style, 2015.
- Com a técnica de transferência de estilo podemos criar nossas próprias obras de arte.
- Transferência de estilo consiste em aplicar o estilo de uma imagem (imagem de referência ou de estilo) em outra imagem (imagem objetivo ou de conteúdo) conservando o conteúdo da imagem objetivo.
- A imagem objetivo é combinada com a imagem de referência gerando uma nova imagem que é uma mistura das duas, mas preserva o conteúdo da imagem objetivo.

O que é transferência de estilo

- Estilo \Rightarrow consiste de textura, cores e padrões visuais presentes na imagem em diferentes escalas espaciais.
- Conteúdo \Rightarrow consiste na macro estrutura da imagem.
- A ideia de transferência de estilo está relacionada com geração de textura, que é uma técnica antiga utilizada na área de processamento de imagens.
- A utilização de RNAs deep learning para transferência de estilo fornece resultados incríveis.

O que é transferência de estilo



Imagem de conteúdo

+



Imagem de estilo

=



Imagem gerada



Imagem de conteúdo

+



Imagem de estilo

=



Imagem gerada

Ativações das camadas intermediárias

- As camadas intermediárias de uma RNA convolucional profunda aprendem padrões existentes nas imagens utilizadas no treinamento.
- Ativações das camadas iniciais de uma RNA profunda contém informações locais da imagem.
- Quanto mais profunda uma camada, mais informação abstrata e global ela representa.
- Quanto mais profunda a camada na RNA mais complexos os padrões aprendidos.
- As ativações das diferentes camadas de uma RNA convolucional profunda fornecem uma decomposição de conteúdos da imagem em escalas espaciais diferentes.

Ativações das camadas intermediárias

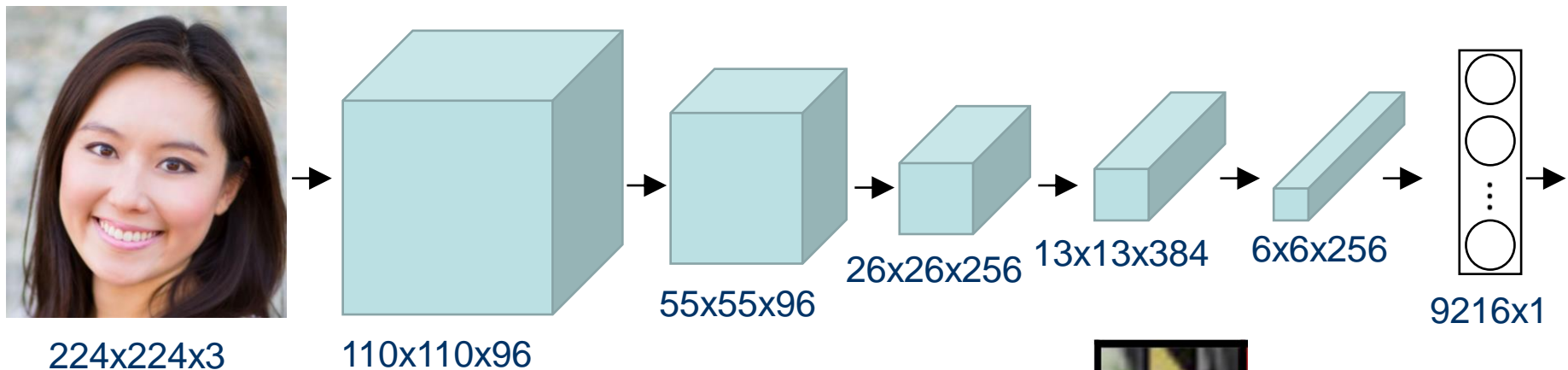
- O conteúdo de uma imagem é uma informação global e abstrata, portanto, espera-se que seja capturado pelas camadas mais profundas.
- Uma forma de visualizar os padrões aprendidos em cada camada de uma RNA convolucional profunda é apresentada por Zeiler and Fergus, “Visualizing and understanding convolutional networks”, 2013.

Ativações das camadas intermediárias

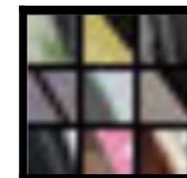
- Método de visualização de Zeiler and Fergus:
 1. Dada uma RNA convolucional, de preferência uma com muitas camadas e treinada com um grande banco de imagens;
 2. Executar essa RNA convolucional com os exemplos de treinamento;
 3. Para cada camada da RNA, determinar as 9 imagens que maximizam as ativações de um determinado filtro;
 4. Fazer o gráfico da saída do filtro desejado para essas 9 imagens;
 5. Repetir esse processo (etapas 3 e 4) para todos os filtros da camada e para as camadas desejadas.

Ativações das camadas intermediárias

- RNA convolucional:



- Dado um filtro da 1ª camada, pegar as 9 imagens que maximizam as ativações desse filtro.
- Repetir para os outros filtros da camada.



Resultado do primeiro filtro

Ativações das camadas intermediárias

- O processo é repetido para outras camadas:
 - Filtros de camadas mais profundas “enxergam” porções maiores das imagens.

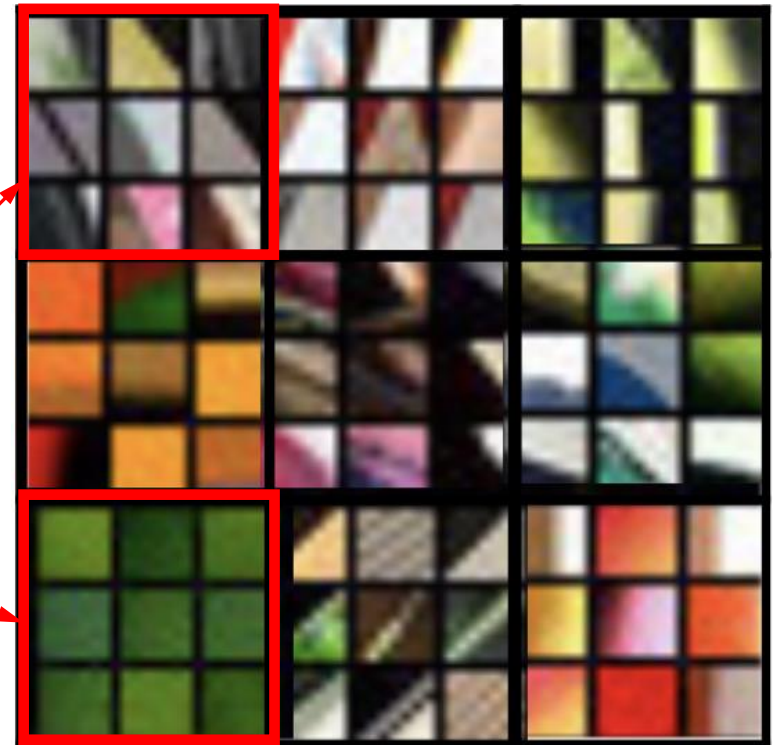


Ativações das camadas intermediárias

- Camada 1:
 - Filtros detectam características básicas \Rightarrow bordas e cores;

Borda inclinada

Cor verde

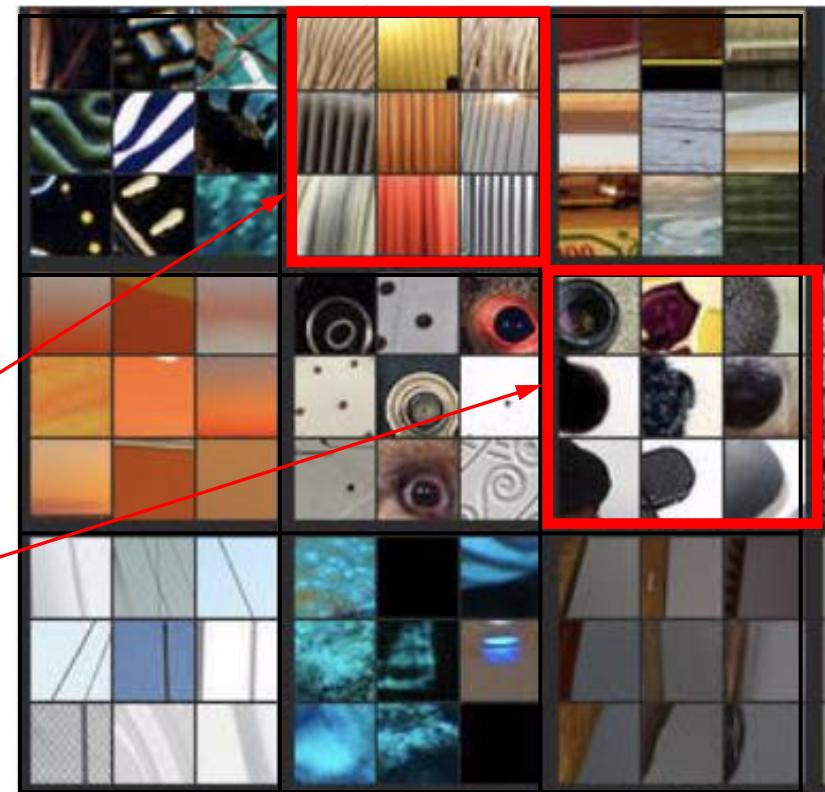


Ativações das camadas intermediárias

- Camada 2:
 - Filtros detectam formas mais complexas \Rightarrow conjunto de linhas, formatos, texturas etc;

Linhas verticais

Formas arredondadas



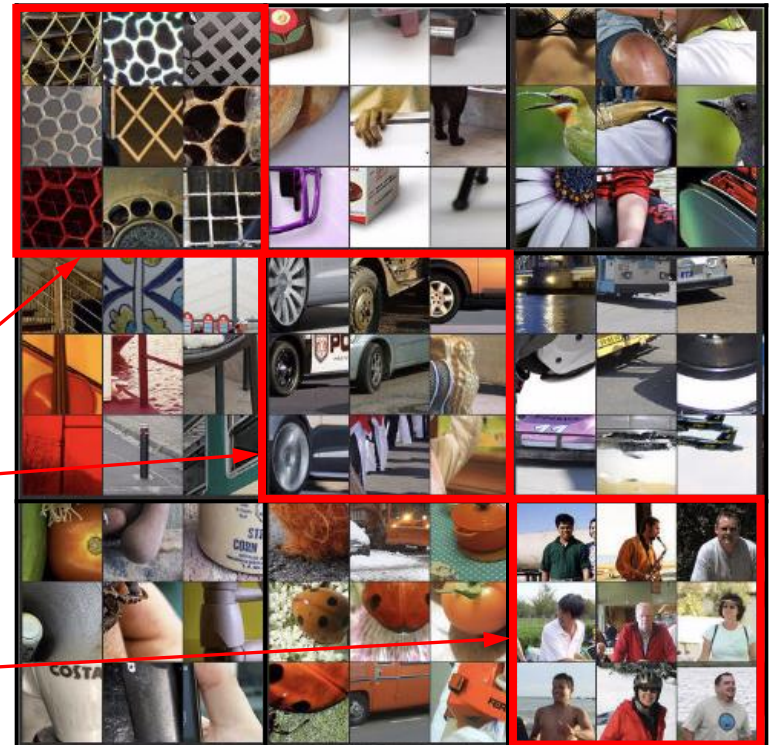
Ativações das camadas intermediárias

- Camada 3:
 - Filtros detectam formas mais complexas \Rightarrow conjunto de linhas, formatos, texturas etc;

Padrões de linhas

Rodas

Pessoas

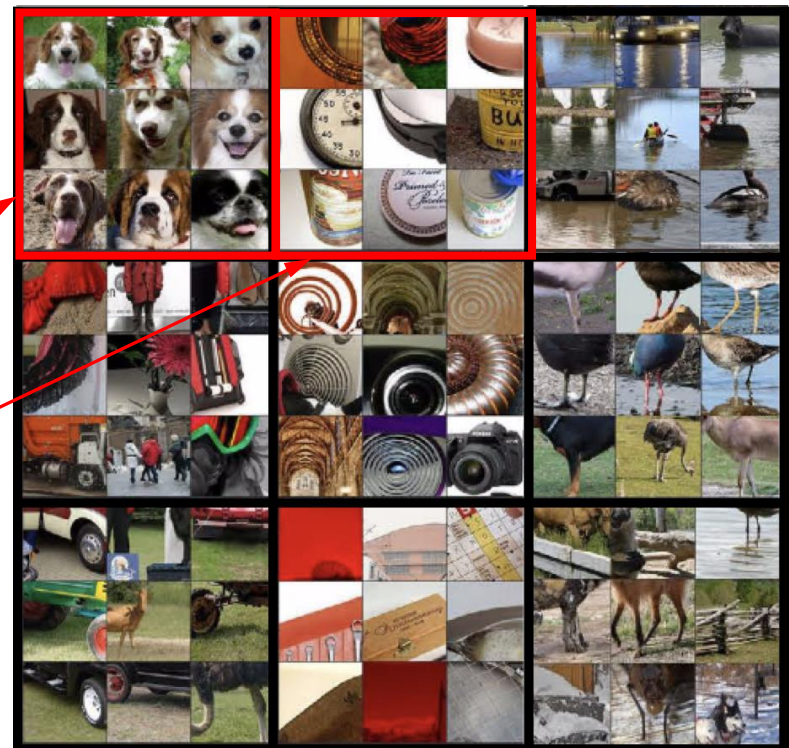


Ativações das camadas intermediárias

- Camada 4:
 - Filtros detectam formas ainda mais complexas \Rightarrow animais, carros, objetos etc;

Cachorros

Objetos



Função de custo

- A ideia básica por trás da transferência de estilo é a mesma utilizada para treinar as redes neurais \Rightarrow define-se uma função de custo para especificar o que deseja-se alcançar e utiliza-se um algoritmo de otimização para minimizar essa função de custo.
- **O que se deseja alcançar** \Rightarrow manter o conteúdo da imagem original mas com o estilo da imagem de referência.
- Se for possível definir matematicamente o que é conteúdo e estilo, então, pode-se definir uma função de custo apropriada.

Função de custo

- A função de custo utilizada é a composição de uma função de custo de conteúdo com uma função de custo de estilo:

$$J(\mathbf{G}) = \alpha J_C(\mathbf{C}, \mathbf{G}) + \beta J_S(\mathbf{S}, \mathbf{G})$$

onde:

$J(\mathbf{G})$ = função de custo;

$J_C(\mathbf{C}, \mathbf{G})$ = parcela da função de custo referente ao conteúdo;

$J_S(\mathbf{S}, \mathbf{G})$ = parcela da função de custo referente ao estilo;

\mathbf{C} = imagem de conteúdo;

\mathbf{S} = imagem de estilo;

\mathbf{G} = nova imagem gerada;

α e β = fatores de ponderação entre as duas funções de custo (hiper-parâmetro a ser definido).

Função de custo

- Parcela de conteúdo da função de custo, $J_C(\mathbf{C}, \mathbf{G})$, mede quão similar é a imagem gerada (\mathbf{G}) em relação à imagem de conteúdo (\mathbf{C}).
- Parcela de estilo da função de custo, $J_S(\mathbf{S}, \mathbf{G})$, mede quão similar é o estilo da imagem gerada (\mathbf{G}) em relação à imagem de estilo (\mathbf{S}).

Geração da nova imagem

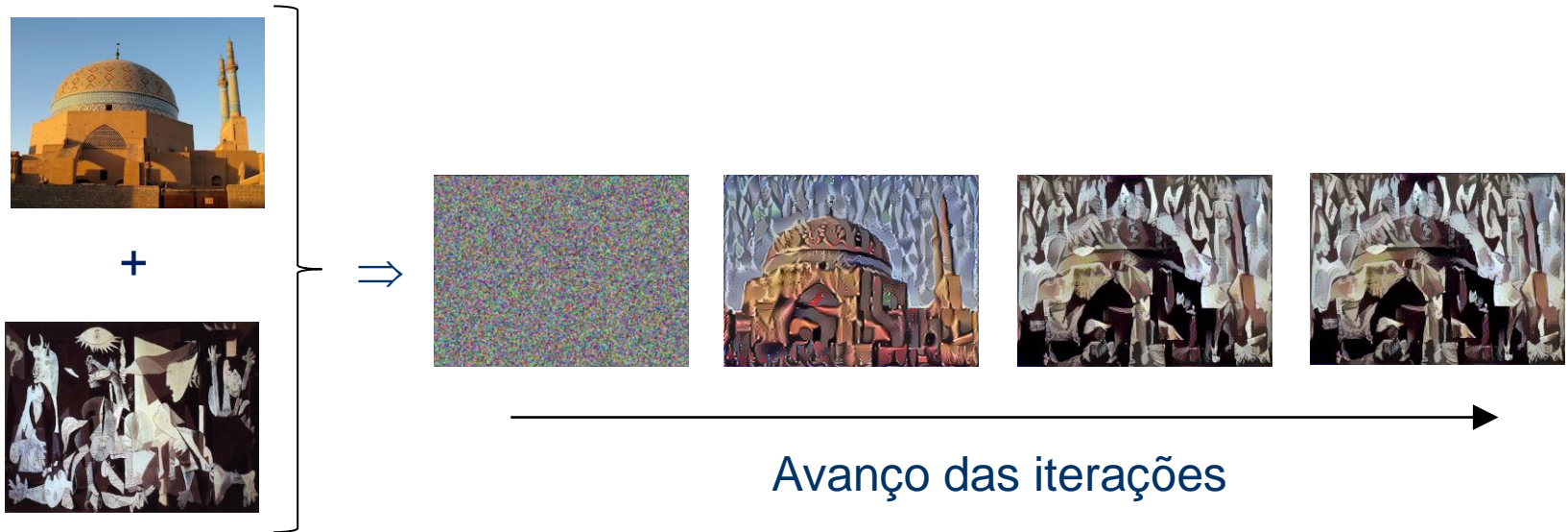
- Geração da imagem **G** é realizada iterativamente.
- Processo de geração da imagem **G**:
 1. Inicializar aleatoriamente a imagem **G** com a dimensão desejada (imagem colorida no formato RGB);
 2. Definir a função de custo e as suas duas parcelas;
 3. Usar o método do Gradiente Descendente para minimizar a função de custo e criar a nova imagem:

$$\mathbf{G} = \mathbf{G} - \frac{\partial J(\mathbf{G})}{\partial \mathbf{G}}$$

O gradiente $\partial J(\mathbf{G})/\partial \mathbf{G}$ é calculado para cada pixel da imagem.

Geração da nova imagem

- Exemplo de geração de uma imagem **G**:



Função de custo de conteúdo

- Para gerar a função de custo de conteúdo usa-se uma RNA convolucional pré-treinada, como por exemplo, a VGG16.
- A imagem de conteúdo (**C**) e a imagem gerada (**G**) são processadas pela RNA pré-treinada.
- Escolhem-se as ativações de uma das camadas para calcular a função de custo.
- Escolhida a camada l para calcular $J_C(\mathbf{C}, \mathbf{G})$, tem-se:
 $\mathbf{a}^{[l]\mathbf{C}}$ = ativações da camada l referentes à imagem conteúdo (**C**);
 $\mathbf{a}^{[l]\mathbf{G}}$ = ativações da camada l referentes à imagem gerada (**G**).
- Se $\mathbf{a}^{[l]\mathbf{C}}$ e $\mathbf{a}^{[l]\mathbf{G}}$ forem similares então as imagens de conteúdo (**C**) e gerada (**G**) são similares.

Função de custo de conteúdo

- Função de custo de conteúdo:

$$J_C(\mathbf{C}, \mathbf{G}) = \frac{1}{2} \|\mathbf{a}^{[l]\mathbf{C}} - \mathbf{a}^{[l]\mathbf{G}}\|_2^2$$

onde $\|\cdot\|_2^2$ representa a soma dos quadrados de cada elemento do vetor ou matriz.

$J_C(\mathbf{C}, \mathbf{G}) \Rightarrow$ representa o erro quadrático médio entre $\mathbf{a}^{[l]\mathbf{C}}$ e $\mathbf{a}^{[l]\mathbf{G}}$.

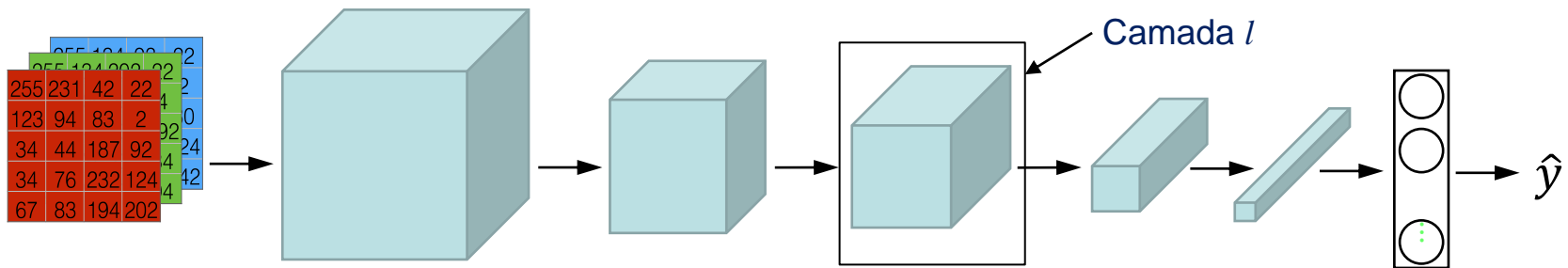
- Na tentativa de minimizar $J_C(\mathbf{C}, \mathbf{G})$ faz-se com que a imagem gerada (\mathbf{G}) se torne similar à imagem de conteúdo (\mathbf{C}).

Função de custo de conteúdo

- Como escolher a camada l adequada:
 - Se usar ativações das camadas iniciais \Rightarrow força a imagem gerada (**G**) ser muito similar à imagem de conteúdo (**C**);
 - Se usar ativações das camadas mais profundas \Rightarrow força a imagem gerada (**G**) ser mais similar à imagem de estilo (**S**);
 - Na prática usam-se as ativações de uma camada intermediária da RNA.

Função de custo de estilo

- O que significa estilo de uma imagem?
 - Para definir o estilo de uma imagem é necessário processar essa imagem por uma RNA convolucional profunda prétreinada;



- Escolhe-se as ativações de uma camada l para medir o “estilo”;
- Estilo é definido como sendo uma correlação entre as ativações dos diversos canais (filtros) da camada;

➤ Porque a correlação entre canais captura estilo?

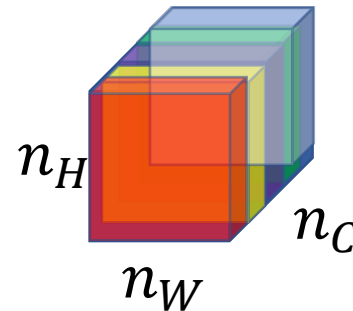
Função de custo de estilo

- Correlação entre as ativações de dois canais de uma camada l :
 - Lembre que cada canal é um filtro que detecta um tipo diferente de característica presente na imagem;
 - **Dois canais estão correlacionados** quando uma imagem ativa muito esses dois canais simultaneamente;
 - **Dois canais não estão correlacionados** quando uma imagem ativa muito um canal e não ativa o outro;
 - Correlação entre dois canais fornece uma medida de quando dois padrões ocorrem simultaneamente em uma dada camada;
 - Grau de correlação entre as ativações de dois canais (padrão) fornece a frequência com que esses dois padrões ocorrem simultaneamente (ou não) em uma imagem ou em partes de uma imagem;
 - Usar correlação entre as ativações dos canais de uma dada camada para medir estilo, permite comparar o estilo da imagem gerada (**G**) com o estilo da imagem de estilo (**S**).

Função de custo de estilo

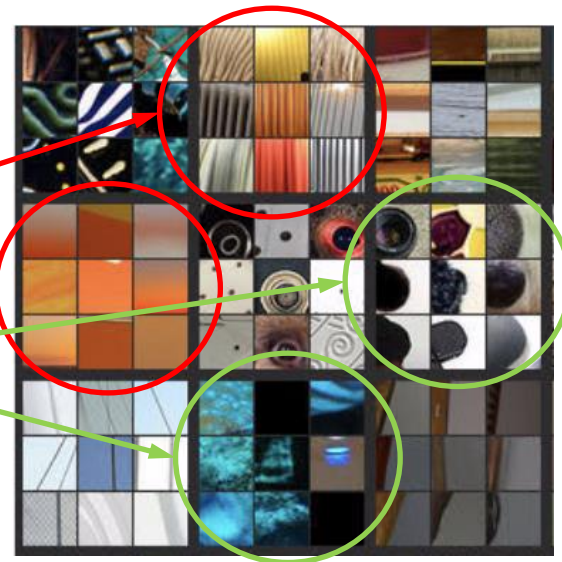
- Correlação entre as ativações de dois canais de uma camada l :

- Camada l com n_c canais \Rightarrow



- Ativações dos n_c canais:

- Correlacionadas
- Não correlacionadas



Função de custo de estilo

- Matriz de estilo é definida pela **Matriz de Gram**.
- Elemento (k, m) da camada l da Matriz de Gram:

$$\mathbf{M}_{k,m}^{[l]} = \sum_{i=1}^{n_H^{[l]}} \sum_{j=1}^{n_W^{[l]}} a_{i,j,k}^{[l]} a_{i,j,m}^{[l]}, \text{ para } k \text{ e } m = 1, \dots, n_c^{[l]}$$

$a_{i,j,k}^{[l]}$ = ativação do pixel i (índice de altura), j (índice de largura), do canal k ;

$a_{i,j,m}^{[l]}$ = ativação do pixel i (índice de altura), j (índice de largura), do canal m .

- Matriz $\mathbf{M}^{[l]}$ possui dimensão $n_c^{[l]} \times n_c^{[l]}$ (matriz quadrada).

Função de custo de estilo

- Elemento $\mathbf{M}_{k,m}^{[l]}$ da Matriz de Gram representa a correlação entre as ativações do canal k com as ativações do canal m da camada l :
 - Se $a_{i,j,k}^{[l]}$ e $a_{i,j,m}^{[l]}$ forem ambos grandes \Rightarrow então $\mathbf{M}_{k,m}^{[l]}$ é grande, então, canais k e m estão correlacionados;
 - Se $a_{i,j,k}^{[l]}$ ou $a_{i,j,m}^{[l]}$ for pequeno \Rightarrow então $\mathbf{M}_{k,m}^{[l]}$ é pequeno, então, canais k e m não estão correlacionados.
- Necessário calcular matrizes de estilos das imagens gerada e de estilo:

$$\mathbf{M}_{k,m}^{[l]G} \text{ e } \mathbf{M}_{k,m}^{[l]S}, \text{ para } k \text{ e } m = 1, \dots, n_c^{[l]}$$

Função de custo de estilo

- Função de custo de estilo para a camada l :

$$J_S^{[l]}(\mathbf{S}, \mathbf{G}) = \frac{1}{4 (n_H^{[l]} n_W^{[l]} n_C^{[l]})^2} \|\mathbf{M}^{[l]} \mathbf{S} - \mathbf{M}^{[l]} \mathbf{G}\|_2^2$$

ou mais detalhadamente:

$$J_S^{[l]}(\mathbf{S}, \mathbf{G}) = \frac{1}{2 (n_H^{[l]} n_W^{[l]} n_C^{[l]})^2} \sum_{m=1}^{n_C^{[l]}} \sum_{k=1}^{n_C^{[l]}} (\mathbf{M}_{k,m}^{[l]} \mathbf{S} - \mathbf{M}_{k,m}^{[l]} \mathbf{G})^2$$

- Usa-se o custo do estilo de várias camadas para formar o custo total de estilo:

$$J_S(\mathbf{S}, \mathbf{G}) = \sum_l \lambda^{[l]} J_S^{[l]}(\mathbf{S}, \mathbf{G})$$

onde $\lambda^{[l]}$ é um hiperparâmetro de ponderação.

Função de custo de estilo

- Uso do custo de estilo de várias camadas permite comparar estilos de características de “baixo” nível (presentes nas primeiras camadas) e de “alto” nível (presentes nas camadas mais profundas).
- A escolha das camadas utilizadas altera o resultado final e depende do que é desejado como imagem gerada.
- Função de custo total é uma soma ponderada das funções de custo de conteúdo e de estilo:

$$J(\mathbf{G}) = \alpha J_C(\mathbf{C}, \mathbf{G}) + \beta J_S(\mathbf{S}, \mathbf{G})$$

- Obter a nova imagem pela minimização de $J(\mathbf{G})$ fornece a imagem com o estilo desejado.

Implementação com TensorFlow e Keras

- Diversos autores implementaram esse método de transferência de estilo usando o TensorFlow e o Keras.
- Todas as funções do método devem ser definidas no programa.
- No trabalho de transferência de estilo veremos como fazer na prática esse processo.