



UNIDADE II

Processamento de Imagem e Visão Computacional

Profa. Dra. Miryam de Moraes

Processamento de imagem e visão computacional

- Operações morfológicas;
- Extração de características;
- Reconhecimento de padrões;
- Visão computacional.

Operações morfológicas

- As operações morfológicas são técnicas fundamentais no processamento de imagens que envolvem a transformação geométrica de objetos em imagens binárias (preto e branco) com base em suas formas e estruturas. Essas operações são utilizadas para aprimorar e analisar características específicas em imagens, especialmente em aplicações de visão computacional e processamento de imagens. As duas operações morfológicas mais comuns são a erosão e a dilatação, frequentemente seguidas pela abertura e pelo fechamento.
- Um elemento estruturante é uma parte fundamental das operações morfológicas no processamento de imagens. Ele é uma pequena matriz ou janela, geralmente de forma geométrica simples.
- O elemento estruturante pode assumir várias formas, como quadrados, retângulos, discos, cruzes.
 - O tamanho do elemento estruturante é determinado pelas dimensões da matriz ou janela.
 - Em muitos casos, o pixel central é posicionado no centro do elemento estruturante, mas, em algumas aplicações, ele pode ser deslocado para outros locais.

Erosão

- A operação de erosão é usada para diminuir o tamanho dos objetos brancos (pixels brancos) em uma imagem binária. Consiste em deslocar um elemento estruturante (também conhecido como kernel) pela imagem e verificar se todos os pixels sob o kernel são brancos. Se todos forem brancos, o pixel central sob o kernel permanece branco; caso contrário, ele é definido como preto. Isso resulta na diminuição do tamanho dos objetos brancos na imagem.

Imagem Original



Imagem Erodida
(Kernel 5)



Imagem Erodida
(Kernel 9)



Dilatação

- A operação de dilatação é usada para aumentar o tamanho dos objetos brancos em uma imagem binária. Assim como na erosão, um kernel é usado, mas, nesse caso, se pelo menos um pixel sob o kernel for branco, o pixel central sob o kernel também é definido como branco.

Imagem Original



Imagem Dilatada
(Kernel 5)



Imagem Dilatada
(Kernel 9)

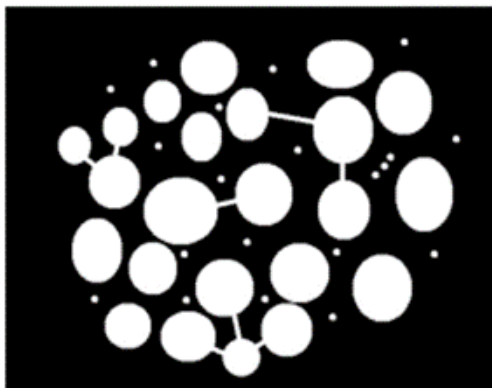


Abertura e fechamento

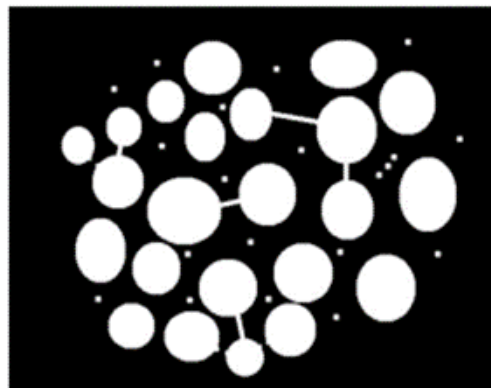
- A abertura é uma operação morfológica que consiste em aplicar a erosão seguida da dilatação em uma imagem. Ela é usada principalmente para remover pequenos detalhes (ruído) e abrir pequenos espaços entre objetos. A abertura pode ser útil para separar objetos que estão muito próximos uns dos outros.
- O fechamento é a operação morfológica oposta à abertura. Ela consiste em aplicar a dilatação seguida da erosão em uma imagem. O fechamento é usado para preencher pequenos buracos em objetos e fechar pequenas lacunas entre eles. Isso pode ser útil para conectar objetos que estão quase se tocando.

Abertura e fechamento

Imagem Original



**Imagem Abertura
(Kernel 5)**

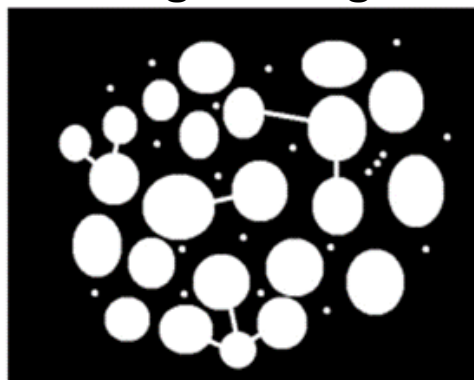


**Imagem Abertura
(Kernel 9)**

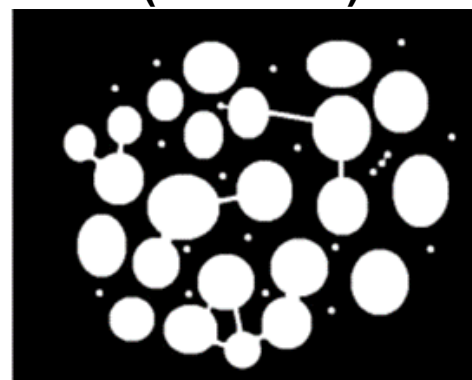


Abertura

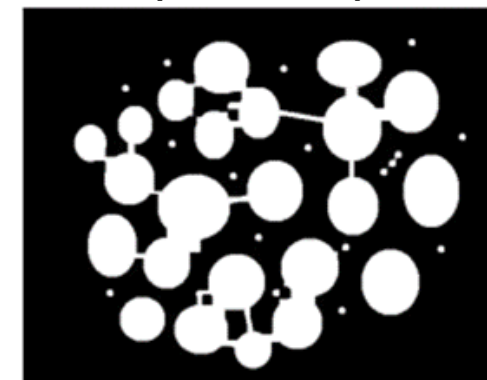
Imagem Original



**Imagem Fechamento
(Kernel 5)**



**Imagem Fechamento
(Kernel 9)**



Fechamento

Gradiente morfológico

- O gradiente morfológico é uma operação morfológica que calcula a diferença entre a dilatação e a erosão de uma imagem. Essa operação é usada para realçar as bordas dos objetos em uma imagem binária. O resultado do gradiente morfológico mostra onde as transições abruptas de valores de pixel ocorrem, o que geralmente corresponde às bordas dos objetos na imagem.

Imagem Original

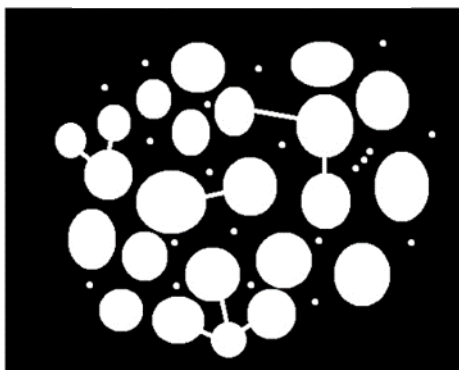


Imagem após Erosão (Kernel 5)

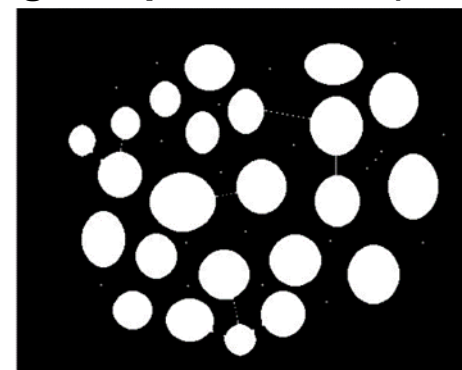
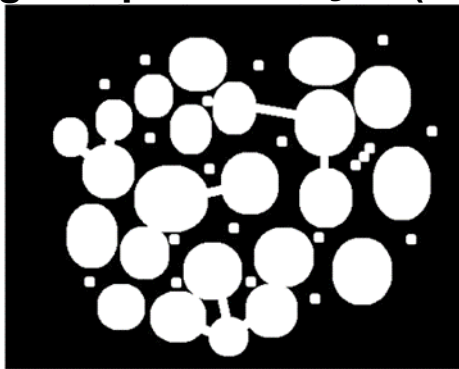
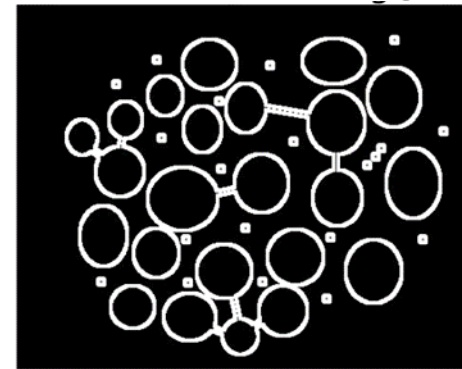


Imagem após Dilatação (Kernel 5)



Gradiente Morfológico



Top hat

- É aplicado em imagens para destacar características que são significativamente menores do que o elemento estruturante usado na operação.
 - Primeiro, é aplicada uma abertura morfológica na imagem original.
 - A imagem original é subtraída do resultado da abertura.
 - O resultado dessa operação é uma imagem que realça as características finas e pequenas da imagem original, como pequenos objetos ou detalhes.

Imagem Original

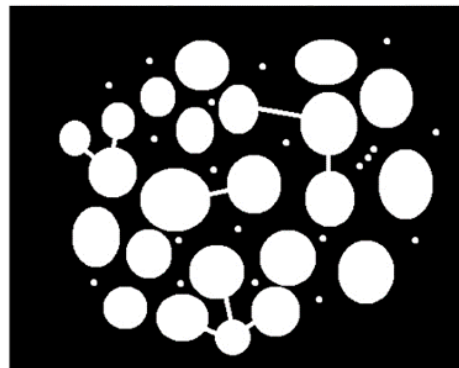
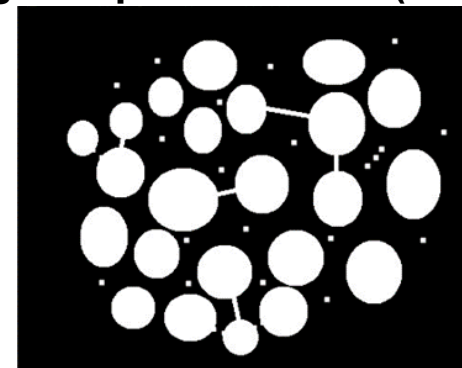
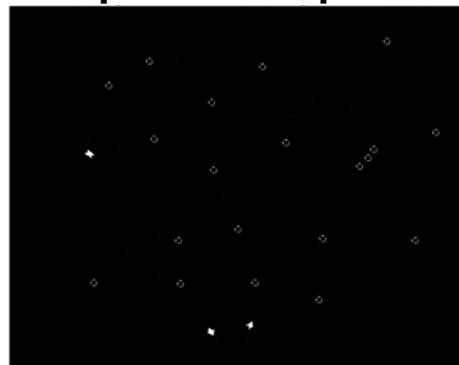


Imagem Após Abertura (Kernel 5)



Operador Top Hat



Detecção de descontinuidades

- As técnicas de detecção de descontinuidade são usadas para identificar transições abruptas de intensidade, que geralmente correspondem a bordas, limites ou descontinuidades em uma imagem.
- Técnicas de detecção de descontinuidade:
- Operadores de gradiente: Sobel, Prewitt e Roberts. Calculam a magnitude do gradiente da intensidade da imagem em direções horizontal e vertical.
 - Operador Laplaciano: calcula a segunda derivada da intensidade da imagem em relação às coordenadas x e y. Regiões com valores de Laplaciano significativamente diferentes de zero são consideradas bordas.
 - Filtro de máscara de desaguçamento (LoG - Laplacian of Gaussian): operador Laplaciano e um filtro Gaussiano. Ele suaviza a imagem com um filtro Gaussiano e, em seguida, calcula o Laplaciano da imagem suavizada para identificar bordas.

Detecção de similaridades

- A segmentação pode seguir uma estratégia de similaridade, em que a partição é efetuada com base na similaridade entre pixels, seguindo um determinado critério (ex: binarização, crescimento de regiões, divisão e junção de regiões).
- A técnica de binarização é comumente usada em visão computacional para destacar objetos de interesse em uma imagem, separando-os do fundo com base na intensidade do pixel.

Imagem Original

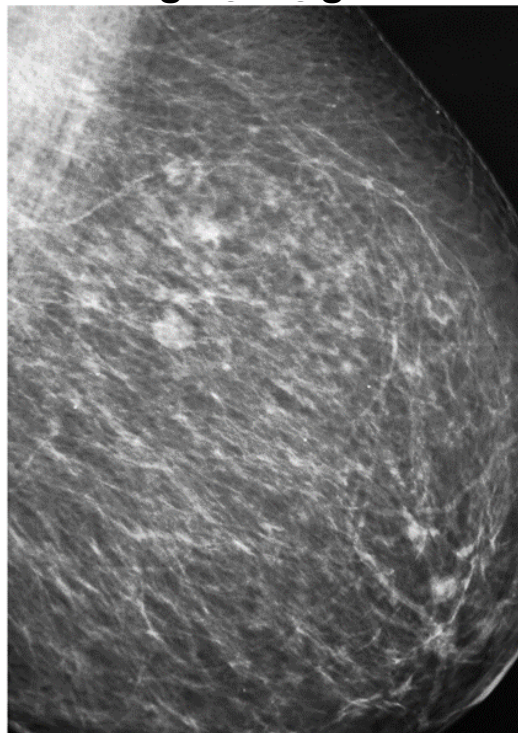
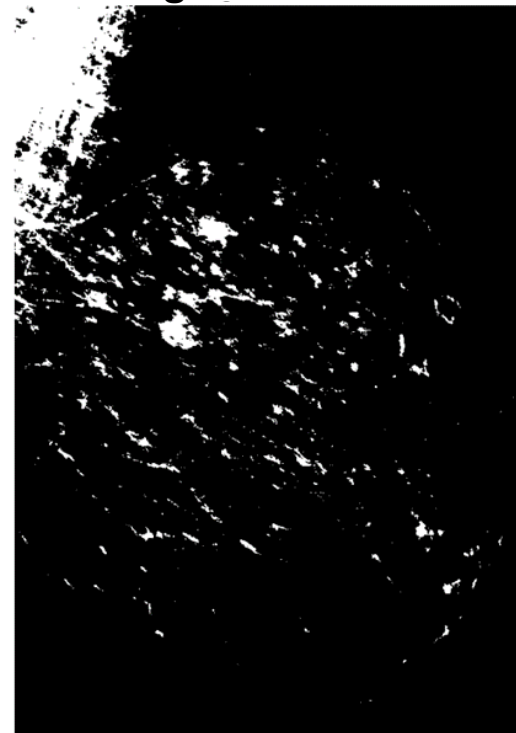


Imagem Binarizada



Mamografia em projeção OML mostrando nódulo de densidade semelhante à do parênquima glandular, com contornos bem definidos, são características de lesão

Fonte:

https://commons.wikimedia.org/wiki/File:N%C3%B3dulo_mamografico_BR0.JPG

Detecção de similaridades: Técnicas

- Correlação cruzada: compara duas imagens deslizando uma sobre a outra e calculando a medida de similaridade entre elas. A posição com a maior medida de similaridade indica uma correspondência.
- Template matching: envolve o uso de uma imagem de referência (template) que é comparada com regiões da imagem de entrada. A região com a melhor correspondência ao template é identificada como uma correspondência.
- Descritores de características: como SIFT (Scale-Invariant Feature Transform) e SURF (Speeded-Up Robust Features), identificam pontos de interesse e descrevem suas características distintas. A correspondência é estabelecida comparando esses descritores entre imagens.
 - Histograma de cores: a comparação de histogramas de cores é usada para medir a similaridade de distribuições de cores em imagens. É útil em aplicações de classificação e recuperação de imagens.

Detecção de similaridades: Técnicas

- Métodos baseados em textura: comparam características de textura, como o padrão de co-ocorrência de intensidade (GLCM) e características de transformada de Fourier para identificar áreas de textura semelhante.
- Correspondência de ponto a ponto: algoritmos de correspondência de ponto a ponto, como o algoritmo Ransac (Random Sample Consensus), são usados para identificar correspondências entre pontos de interesse em diferentes imagens.
- Descritores locais: são usados, como o descritor HOG (Histogram of Oriented Gradients), para identificar objetos com base nas características locais, como bordas e gradientes.
 - Aprendizado de máquina: redes neurais convolucionais (CNNs), podem ser treinados para realizar tarefas de correspondência de objetos e reconhecimento de padrões.

Interatividade

Considere as seguintes asserções:

- I. O tamanho afeta diretamente o resultado da operação morfológica. Elementos maiores tendem a fazer com que as operações dilatem ou erodam mais, enquanto elementos menores têm um efeito mais sutil.
- II. Pode assumir várias formas, como quadrados, retângulos, discos, cruzes, entre outros.
- III. É uma pequena matriz ou janela, geralmente de forma geométrica simples, usada para aplicar transformações morfológicas.

Tais asserções referem-se a:

- a) Filtros de frequência.
- b) Elementos estruturantes.
- c) Filtros espaciais.
- d) Erosões.
- e) Dilatações.

Resposta

Considere as seguintes asserções:

- I. O tamanho afeta diretamente o resultado da operação morfológica. Elementos maiores tendem a fazer com que as operações dilatem ou erodam mais, enquanto elementos menores têm um efeito mais sutil.
- II. Pode assumir várias formas, como quadrados, retângulos, discos, cruzes, entre outros.
- III. É uma pequena matriz ou janela, geralmente de forma geométrica simples, usada para aplicar transformações morfológicas.

Tais asserções referem-se a:

- a) Filtros de frequência.
- b) Elementos estruturantes.**
- c) Filtros espaciais.
- d) Erosões.
- e) Dilatações.

Extração de características

- A extração de características em processamento de imagem refere-se ao processo de identificar, isolar e medir características específicas em uma imagem que são relevantes para uma determinada tarefa de análise ou reconhecimento.
- Essas características podem ser estruturais, estatísticas, texturais, geométricas ou baseadas em intensidade, dependendo dos requisitos da aplicação.
- Após a extração, as características podem ser usadas para treinar algoritmos de aprendizado de máquina, realizar correspondência de padrões, segmentar objetos, reconhecer objetos.
 - A redução de dimensionalidade em processamento de imagem é usada em conjunto com técnicas de extração de características para melhorar a eficiência e a eficácia da análise de imagens.
 - Redução de dimensionalidade linear e não linear.

Maldição da dimensionalidade

- A maldição da dimensionalidade refere-se aos desafios e complicações que surgem quando se lida com conjuntos de dados de alta dimensionalidade, em que o número de características ou dimensões é significativamente maior do que o número de amostras de dados disponíveis.
- À medida que adicionamos mais dimensões aos dados, precisamos de uma quantidade exponencialmente maior de dados para representar adequadamente essas dimensões.

Estratégias adotadas:

- Seleção de características: identificar e manter apenas as características mais relevantes e informativas, eliminando as redundantes ou irrelevantes.
 - Extração de características: usar técnicas de extração de características para reduzir a dimensionalidade, transformando os dados em um espaço de características mais compacto e informativo.

Maldição da dimensionalidade

Estratégias (continuação):

- Agrupamento e redução de dimensionalidade: aplicar técnicas de agrupamento para agrupar características relacionadas e reduzir a dimensionalidade de forma controlada.
- Aumento de dados: se possível, coletar mais dados para lidar com o problema de esparsidade de dados em dimensões elevadas.
 - Visualização de dados: quando possível, usar técnicas de visualização, como PCA ou t-SNE, para projetar os dados em um espaço de menor dimensão para uma melhor compreensão e interpretação.

Segmentação no espaço de atributos

- Espaço de atributos: o espaço de atributos é uma representação multidimensional que descreve as características dos pixels em uma imagem. Cada dimensão do espaço de atributos corresponde a uma característica específica, como intensidade, cor, textura, gradiente, entre outros.
- A segmentação no espaço de atributos envolve a divisão de uma imagem em regiões ou objetos com base nas características dos pixels. Essas características podem ser extraídas do espaço de atributos da imagem.
 - Segmentação por limiarização é uma técnica simples de segmentação que se baseia na definição de um ou mais limiares no espaço de atributos. Pixels com valores de atributo que atendem aos critérios de limiar são atribuídos a uma classe ou região específica. Por exemplo, na limiarização de intensidade, pixels com intensidade acima de um limiar são considerados como objeto, enquanto aqueles abaixo do limiar são considerados como fundo.

Segmentação no espaço de atributos

- Segmentação por regiões: em vez de usar limiares fixos, a segmentação por regiões busca identificar grupos de pixels com características semelhantes no espaço de atributos. Isso é frequentemente feito usando técnicas de agrupamento, como k-means ou regiões crescentes. Os pixels são agrupados em regiões com base na proximidade das características no espaço de atributos.
- Segmentação por borda: a segmentação por borda visa identificar as bordas ou contornos dos objetos na imagem. Isso é feito detectando transições abruptas nas características, como mudanças na intensidade, gradiente ou textura. Algoritmos de detecção de borda, como o operador Sobel ou Canny, são comumente usados para essa finalidade.
 - Segmentação por crescimento de região: esse método começa com um ou mais pontos de semente na imagem e expande uma região conectada em torno dessas sementes com base em critérios de similaridade no espaço de atributos. Isso é útil para segmentar objetos conectados ou áreas homogêneas.

Segmentação no espaço de atributos

- Segmentação por aprendizado de máquina: O aprendizado de máquina, como em redes neurais convolucionais (CNNs), também é utilizado para a segmentação no espaço de atributos. Esses modelos podem aprender automaticamente a identificar objetos com base em um grande conjunto de dados de treinamento.
- Segmentação semântica: é uma forma avançada de segmentação que atribui uma classe semântica a cada pixel na imagem. Isso é comum em aplicações de visão computacional, como veículos autônomos e processamento de imagens médicas.
 - Muitas vezes, uma combinação de técnicas é usada para obter resultados mais precisos e robustos.

Análise de Componentes Principais (PCA)

- A PCA é uma técnica de transformação linear que calcula as direções (componentes principais) ao longo das quais os dados têm a maior variabilidade.
- Essas direções correspondem aos autovetores da matriz de covariância dos dados. Ao projetar os dados nas direções dos autovetores mais importantes, é possível reduzir a dimensionalidade dos dados, mantendo uma grande parte da informação original.
- Coleta de dados: as imagens são coletadas e representadas como matrizes de pixels, em que cada pixel é uma dimensão.
- Padronização: os dados são geralmente padronizados (subtraindo a média e dividindo pelo desvio padrão) para garantir que todas as dimensões tenham uma escala comparável.
 - Cálculo da matriz de covariância: a matriz de covariância dos dados é calculada. Essa matriz descreve como as diferentes dimensões dos dados estão correlacionadas entre si.

Análise de Componentes Principais (PCA)

- Cálculo dos autovetores e autovalores: os autovetores e autovalores da matriz de covariância são calculados. Os autovetores representam as direções principais nos dados, enquanto os autovalores indicam a importância relativa dessas direções.
- Seleção das componentes principais: as principais componentes são selecionadas com base na magnitude dos autovalores. Normalmente, as componentes principais são ordenadas em ordem decrescente de importância.
 - Projeção dos dados: os dados são projetados nas principais componentes selecionadas. Isso reduz a dimensionalidade dos dados, criando uma representação das imagens em um espaço de dimensão reduzida.

Aplicação da PCA em processamento de imagem

- Redução de dimensionalidade: em imagens digitais, cada pixel pode ser considerado uma dimensão, o que pode resultar em um espaço de alta dimensionalidade. A PCA permite reduzir a dimensionalidade das imagens, mantendo as características mais informativas. Isso é útil para economizar espaço de armazenamento, acelerar algoritmos de processamento e análise de imagem e melhorar a visualização de dados.
- Compressão de imagem: a PCA é usada em técnicas de compressão de imagem, como a compressão de imagem JPEG.
- Análise de texturas: a PCA pode ser aplicada a texturas em imagens para reduzir a dimensionalidade e identificar características importantes das texturas, facilitando a classificação e a análise.
 - Remoção de ruído: a PCA pode ser usada para separar informações de sinal de ruído em imagens, concentrando-se nas principais componentes que representam o sinal.

Análise Discriminante Linear (LDA)

- A Análise Discriminante Linear (LDA) é uma técnica estatística e de aprendizado de máquina usada em processamento de imagem e em várias outras áreas para a redução de dimensionalidade e a classificação de dados. Em processamento de imagem, a LDA é aplicada principalmente para melhorar a eficiência da classificação de imagens, especialmente em tarefas de reconhecimento de padrões.
- A LDA é uma técnica que se concentra na maximização da separação entre diferentes classes de dados, ao mesmo tempo em que minimiza a dispersão dentro das classes.
- Coleta de dados de treinamento: um conjunto de imagens de treinamento é coletado, cada uma rotulada com sua classe correspondente.
 - Extração de características: características relevantes são extraídas das imagens, como intensidades de pixel, texturas, gradientes, entre outras.
 - Padronização: os dados de características são geralmente padronizados para garantir que todas as características tenham uma escala comparável.

Análise Discriminante Linear (LDA)

- Cálculo das médias e matriz de dispersão: a média das características é calculada para cada classe, bem como uma matriz de dispersão dentro da classe e uma matriz de dispersão entre as classes.
- Cálculo dos autovetores e autovalores: a LDA calcula os autovetores e autovalores da matriz resultante da multiplicação da matriz de dispersão entre classes pela matriz de dispersão dentro das classes.
- Seleção das componentes principais: as componentes principais são selecionadas com base na magnitude dos autovalores. Normalmente, as componentes principais são ordenadas em ordem decrescente de importância discriminante.
 - Projeção dos dados: os dados de características são projetados nas principais componentes selecionadas. Isso reduz a dimensionalidade dos dados, criando uma representação das imagens em um espaço de dimensão reduzida, que é otimizado para a discriminação entre classes.

Aplicação da LDA em processamento de imagem

- Classificação de imagens: a LDA é usada para melhorar a classificação de imagens, identificando as características mais discriminantes para separar diferentes classes de objetos ou padrões em imagens. Isso é útil em tarefas como reconhecimento de objetos, detecção de faces, classificação de documentos, entre outras.
- Redução de dimensionalidade supervisionada: a LDA permite reduzir a dimensionalidade dos dados enquanto mantém as informações mais discriminantes. Isso é útil para acelerar algoritmos de processamento, economizar espaço de armazenamento e melhorar a eficácia da classificação.
- Reconhecimento de padrões: ajuda a encontrar as características mais importantes que distinguem diferentes padrões ou classes de objetos em imagens. É crucial.
 - A LDA é uma técnica poderosa quando se trata de classificação de imagens e reconhecimento de padrões.
 - Ajuda a encontrar as características mais discriminantes para melhorar a precisão da classificação e a eficácia da análise de imagem.

Interatividade

Considere as seguintes asserções:

- I. Refere-se aos desafios e complicações que surgem quando se lida com conjuntos de dados, em que o número de características ou dimensões é significativamente maior do que o número de amostras de dados disponíveis.
- II. Ocorre porque à medida que se adicionam mais dimensões aos dados, torna-se necessária uma quantidade exponencialmente maior de dados para representar adequadamente essas dimensões.
- III. Desdobra-se em um aumento de custos computacionais.

Tais asserções estão tratando:

- a) Da maldição das características.
- b) Da maldição da erosão.
- c) Da maldição da dimensionalidade.
- d) Da erosão.
- e) Do overfitting.

Resposta

Considere as seguintes asserções:

- I. Refere-se aos desafios e complicações que surgem quando se lida com conjuntos de dados, em que o número de características ou dimensões é significativamente maior do que o número de amostras de dados disponíveis.
- II. Ocorre porque à medida que se adicionam mais dimensões aos dados, torna-se necessária uma quantidade exponencialmente maior de dados para representar adequadamente essas dimensões.
- III. Desdobra-se em um aumento de custos computacionais.

Tais asserções estão tratando:

- a) Da maldição das características.
- b) Da maldição da erosão.
- c) **Da maldição da dimensionalidade.**
- d) Da erosão.
- e) Do overfitting.

Reconhecimento de padrões

- O reconhecimento de padrões é a tarefa de identificar automaticamente padrões, objetos ou informações em imagens digitais com base em características específicas extraídas das imagens. Isso envolve a análise das características de interesse nas imagens e a comparação dessas características com modelos ou padrões predefinidos para tomar decisões sobre a presença ou ausência de um padrão ou objeto específico



Um exemplo de conjunto de dados de números MNIST

Fonte: <https://commons.wikimedia.org/wiki/File:MnistExamplesModified.png>

Classificador de k vizinhos mais próximos

- Coleta de dados de treinamento: um conjunto de imagens de treinamento é coletado e cada imagem é associada a uma classe (rótulo) correspondente. Essas classes podem representar diferentes objetos, categorias ou padrões.
- Extração de características: a partir das imagens de treinamento, características relevantes são extraídas. No processamento de imagem, isso pode incluir texturas, histogramas de cores, descritores de formas, gradientes, entre outros.
- Padronização dos dados: as características são frequentemente padronizadas (subtraindo a média e dividindo pelo desvio padrão) para garantir que todas tenham uma escala comparável.
 - Treinamento do modelo: os dados de treinamento são usados para treinar o modelo k-NN. O modelo armazena as características das imagens de treinamento e suas classes correspondentes.

Classificador de k vizinhos mais próximos

- Classificação de imagens de teste: para uma nova imagem que precisa ser classificada, as características também são extraídas. O algoritmo k-NN calcula a distância entre as características da imagem de teste e as características das imagens de treinamento. Ele seleciona os k exemplos de treinamento mais próximos (os vizinhos) com base nas menores distâncias.
- Votação de maioria: o algoritmo determina a classe mais frequente entre os k vizinhos mais próximos e atribui essa classe à imagem de teste. Por exemplo, se a maioria dos vizinhos pertencer à classe "cachorro", a imagem de teste será classificada como um "cachorro".

Método de classificação em cascata (Haar Cascade)

- O método Haar Cascade é conhecido por sua eficiência na detecção de objetos em tempo real, como rostos, olhos, carros, placas de veículos e muitos outros. O Haar Cascade é baseado na técnica de classificação em cascata, que envolve a aplicação sequencial de classificadores para filtrar regiões de interesse em uma imagem.

O processo de detecção de objetos usando Haar Cascade pode ser resumido em quatro etapas principais:

- Treinamento do classificador em cascata: a primeira etapa envolve o treinamento do classificador em cascata usando um grande conjunto de dados de treinamento que contém exemplos positivos (imagens que contêm o objeto de interesse) e exemplos negativos (imagens que não contêm o objeto de interesse). O classificador é treinado para distinguir entre as duas classes com base nas características de Haar.

Método de classificação em cascata (Haar Cascade)

- Criação de um classificador em cascata: o classificador em cascata é criado combinando vários classificadores fracos (geralmente classificadores de árvore de decisão simples) em uma cascata. Cada classificador na cascata é treinado para ser rápido e descartar rapidamente as regiões que não contêm o objeto de interesse.
- Detecção em cascata: a imagem de entrada é dividida em várias janelas sobrepostas de diferentes tamanhos. Em cada janela, o classificador em cascata é aplicado sequencialmente. Se uma janela falhar em qualquer estágio da cascata, ela é descartada. Somente se passar por todos os estágios, a janela será considerada uma detecção positiva.
 - Pós-processamento: as detecções positivas são refinadas e agrupadas para formar regiões de objetos. Isso pode envolver a eliminação de sobreposições, o ajuste de retângulos de detecção e a rejeição de detecções fracas.

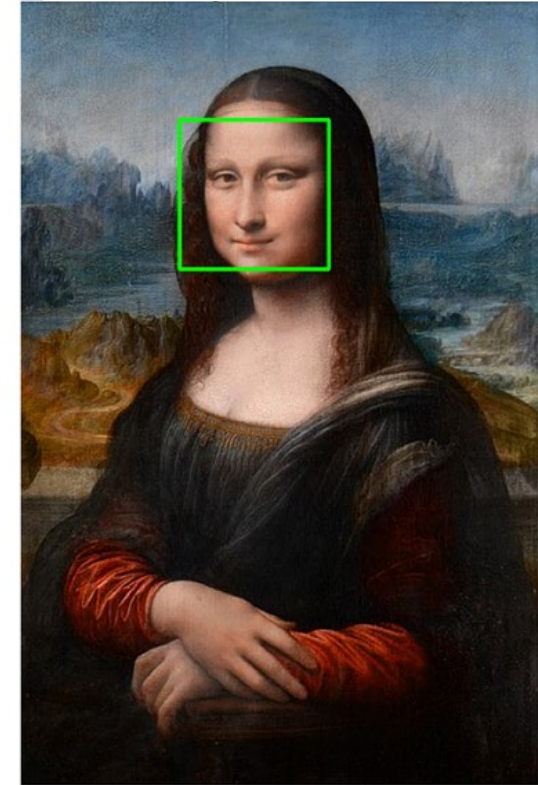
Método de classificação em cascata (Haar Cascade)

- Para a detecção de rosto, pode-se usar o classificador Haar Cascade, que é pré-treinado para detecção de rostos.
- Trata-se de uso do OpenCV.

Imagem Original



Detecção de Rostos



Leonardo da Vinci: Mona Lisa
Fonte: <https://shre.ink/Ux0z>

Estado da arte

- Inception-v4 and ResNet-152 (2016): essas arquiteturas aprimoradas demonstraram o poder das redes profundas e eficientes na classificação de imagens em grandes conjuntos de dados.
- AlphaGo (2016): a equipe do Google DeepMind desenvolveu o AlphaGo, um sistema de IA que derrotou o campeão mundial de Go. Isso demonstrou a capacidade de máquinas de aprender e reconhecer padrões em jogos complexos.
- Transformers e GPT-3 (2018-2020): a arquitetura Transformer revolucionou o processamento de linguagem natural e a visão computacional. O GPT-3, desenvolvido pela OpenAI, é um exemplo notável que demonstrou habilidades impressionantes em reconhecimento de padrões em texto.
 - É importante observar que esses avanços muitas vezes resultam de uma colaboração interdisciplinar entre matemáticos, cientistas da computação, engenheiros e especialistas em domínios específicos, e a evolução continua.

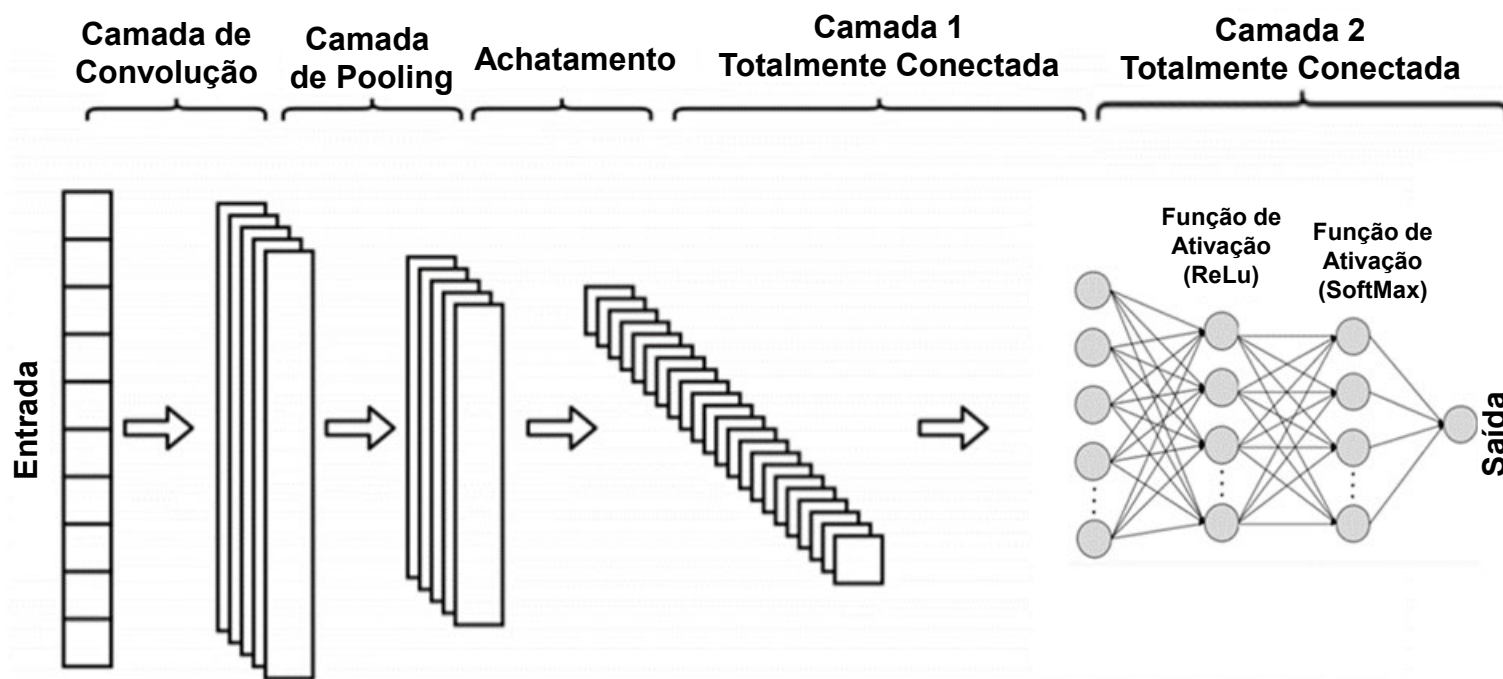
Abordagens baseadas em aprendizado profundo

- Abordagens baseadas em aprendizado profundo (Deep Learning) revolucionaram o campo do reconhecimento de padrões nas últimas décadas. Essas técnicas têm se destacado em várias tarefas de visão computacional, processamento de linguagem natural e muito mais.
- As CNNs são uma classe fundamental de redes neurais profundas que têm sido especialmente eficazes em tarefas de visão computacional, como classificação de imagens, detecção de objetos e segmentação semântica.
- São projetadas para lidar com dados, como imagens, e empregam camadas de convolução para extrair características hierárquicas. O trabalho seminal da equipe do ImageNet em 2012, liderado por Alex Krizhevsky.
 - Desencadeou o uso generalizado de CNNs em tarefas de visão computacional.
 - Aplicações de realidade aumentada e virtual usam aprendizado profundo para rastreamento de marcadores, detecção de planos e detecção de objetos.
 - Importante na percepção de veículos autônomos.

Abordagens baseadas em aprendizado profundo

- Redes neurais, como a Faster R-CNN e a Yolo (You Only Look Once), são usadas para detecção de objetos em imagens, sendo aplicadas em sistemas de vigilância, veículos autônomos e muito mais. Redes neurais profundas, como as redes de segmentação semântica, permitem a identificação precisa de áreas de interesse em imagens, sendo úteis em áreas como medicina, geologia e análise de satélite.
- Redes Generativas Adversariais (GANs) : O trabalho do BigGAN e do DALL-E, da OpenAI, são exemplos notáveis de geração de imagens usando GANs. O aprendizado profundo é usado na classificação de imagens médicas, auxiliando no diagnóstico de doenças a partir de imagens de raios-X, ressonâncias magnéticas e tomografias computadorizadas.
 - O reconhecimento de expressões faciais é uma aplicação popular que utiliza CNNs para identificar emoções em imagens e vídeos. O processamento de imagens em astronomia é aprimorado pelo uso de aprendizado profundo para classificar estrelas, planetas e outros objetos celestes.

Redes neurais artificiais convolucionais



Redes Neurais Convolucionais
Fonte: adaptado de: livro-texto

- Estrutura de camadas: as CNNs consistem em camadas empilhadas, incluindo camadas de convolução, camadas de pooling e camadas totalmente conectadas. As camadas de convolução são o elemento central das CNNs e são responsáveis por extrair características das imagens.

Redes neurais artificiais convolucionais

- Camadas de convolução: as camadas de convolução aplicam filtros (também chamados de kernels) a pequenas regiões da imagem. Esses filtros detectam padrões simples, como bordas, texturas e formas. À medida que você avança nas camadas, os filtros se tornam mais abstratos e capturam características mais complexas.
- Camadas de pooling: as camadas de pooling reduzem a dimensionalidade da imagem, preservando as características mais importantes. Elas fazem isso agrupando informações de regiões próximas e mantendo apenas a informação mais significativa. Isso ajuda a reduzir a quantidade de parâmetros e a evitar o overfitting.
 - Camadas totalmente conectadas: após as camadas de convolução e pooling, as CNNs geralmente incluem camadas totalmente conectadas para realizar a classificação ou outra tarefa específica. Essas camadas são semelhantes às camadas de uma rede neural tradicional.

Redes neurais artificiais convolucionais

- Aprendizado hierárquico de características: uma das principais vantagens das CNNs é sua capacidade de aprender características hierárquicas automaticamente. As camadas iniciais capturam detalhes de baixo nível, enquanto as camadas posteriores combinam essas características para formar representações mais abstratas.

Operação de convolução

- A convolução é uma operação matemática que combina duas funções para criar uma terceira função. No contexto do reconhecimento de padrões e processamento de imagens, a convolução envolve uma matriz chamada "filtro" ou "kernel" que desliza sobre uma matriz de dados de entrada (por exemplo, uma imagem), multiplicando e somando elementos correspondentes.
- Aplicação da convolução em imagens:
- Filtro (kernel): o filtro é uma matriz menor que atua como um detector de características. Por exemplo, um filtro pode ser projetado para detectar bordas horizontais em uma imagem.
 - Imagem de entrada: a imagem é representada como uma matriz de pixels. O filtro é aplicado a pequenas regiões da imagem por vez, começando na parte superior esquerda e movendo-se gradualmente para a direita e para baixo.
 - Operação de multiplicação e soma: em cada etapa, o filtro é sobreposto à região da imagem, e os elementos correspondentes são multiplicados. Os produtos são então somados para produzir um único valor.

Operação de convolução

- Mapa de características (feature map): O resultado da convolução é chamado de "mapa de características". Ele destaca a presença de características específicas, como bordas, texturas ou padrões na imagem.
- Aprendizado de características hierárquicas: a convolução apresenta a capacidade de aprender características hierárquicas automaticamente. Os filtros em camadas mais profundas de uma CNN podem combinar informações de camadas mais rasas para identificar características mais complexas. Por exemplo, filtros nas primeiras camadas podem aprender a detectar bordas, enquanto filtros nas camadas posteriores podem aprender a detectar formas mais complexas que consistem em bordas.
 - Redução da dimensionalidade: as camadas de pooling após as camadas de convolução, resumem informações em regiões maiores, preservando as características mais importantes e reduzindo a quantidade de dados processados nas camadas seguintes.

Operação de convolução

Aplicações no reconhecimento de padrões:

- Reconhecimento de padrões que envolvem imagens, como classificação de imagens, detecção de objetos.
- Segmentação semântica e muito mais. Personalização de filtros.
- Um dos benefícios das CNNs: os filtros são aprendidos durante o treinamento da rede. Isso significa que a rede neural pode adaptar seus filtros para serem específicos para a tarefa em questão, resultando em melhor desempenho.
 - A operação de convolução permite a extração automática de características relevantes de dados de grade, como imagens.

Interatividade

A convolução é uma operação matemática que combina duas funções para criar uma terceira função. Assinale a alternativa que elenca quais são essas três funções:

- a) Erosão, pooling, abertura.
- b) Reconhecimento de padrões, detecção de bordas, segmentação.
- c) Reconhecimento de padrões, detecção de bordas, suavização.
- d) Filtro, imagem de entrada, mapa de características.
- e) Operações de soma e multiplicação.

Resposta

A convolução é uma operação matemática que combina duas funções para criar uma terceira função. Assinale a alternativa que elenca quais são essas três funções:

- a) Erosão, pooling, abertura.
- b) Reconhecimento de padrões, detecção de bordas, segmentação.
- c) Reconhecimento de padrões, detecção de bordas, suavização.
- d) Filtro, imagem de entrada, mapa de características.**
- e) Operações de soma e multiplicação.

Pooling

- Pooling, também conhecido como subamostragem, é uma operação utilizada para reduzir a dimensionalidade dos dados, tornando-os mais gerenciáveis e eficientes computacionalmente.

Os dois tipos mais comuns de pooling são:

- Max Pooling: essa operação seleciona o valor máximo de um grupo de valores em uma região específica da matriz de entrada. O Max Pooling é frequentemente usado para preservar características dominantes em uma região da imagem.
- Average Pooling: calcula a média dos valores em uma região da matriz de entrada. Isso pode ser útil quando se deseja uma representação mais suavizada dos dados.
 - Redução da dimensionalidade, invariância à translação, localização aproximada das características.

Flatten Layer

- A Flatten Layer, ou camada de achatamento, é um componente fundamental em reconhecimento de padrões. Essa camada desempenha um papel importante na transição entre as camadas de convolução e as camadas totalmente conectadas em uma CNN.
- A Flatten Layer é uma camada simples que tem a tarefa de transformar os dados tridimensionais (altura x largura x canais, obtidas pelas camadas de pooling) em um vetor unidimensional destinado à entrada das camadas totalmente conectadas, também conhecidas como camadas densas, em que cada valor é considerado como um recurso independente. Exemplo: se a saída de uma camada de convolução for uma matriz de tamanho $4 \times 4 \times 64$, a Flatten Layer a transformará em um vetor de 1024 elementos ($4 \times 4 \times 64 = 1024$).
 - As camadas totalmente conectadas da rede neural podem conter neurônios que estão conectados a todos os elementos do vetor, permitindo que a rede aprenda relações.

Funções de ativação

- Funções de ativação são essenciais no reconhecimento de padrões, especialmente em redes neurais artificiais e em Deep Learning. Elas são funções matemáticas aplicadas nas saídas das camadas de uma rede neural e são responsáveis por introduzir não linearidades, tornando possível que a rede aprenda representações complexas e não lineares dos dados.
- As funções de ativação determinam se um neurônio artificial deve ser ativado (ou seja, transmitir um sinal adiante) ou não.
 - Importância da não linearidade: permitindo que elas capturem e modelem padrões complexos em dados, o que é crucial para tarefas de reconhecimento de padrões.

Funções de ativação

Funções de ativação mais comuns:

- Sigmoid: mapeia os valores de entrada para um intervalo entre 0 e 1. Foi muito utilizada em redes neurais mais antigas, mas tem caído em desuso em favor de funções mais eficazes.

$$f(x) = 1/(1 + e^{(-x)})$$

- Função de etapa binária: a função de ativação é um classificador baseado em limiar (threshold), isto é, se o neurônio deve ou não ser ativado.

$$f(x) = 1, x \geq 0$$

$$f(x) = 0, x < 0$$

- Tangente hiperbólica (tanh): mapeia os valores de entrada para um intervalo entre -1 e 1, tornando-a mais centrada na origem do que a sigmoid.

$$\tanh(x) = 2/(1 + e^{(-2x)}) - 1$$

Funções de ativação

- ReLU (Rectified Linear Unit): retorna zero para valores negativos e o próprio valor de entrada para valores não negativos. Ela é usada devido à sua simplicidade e eficácia.

$$f(x) = \max(0, x)$$

- Leaky ReLU: é uma variante da função ReLU que permite que uma pequena inclinação seja aplicada à parte negativa dos valores de entrada, evitando o problema do "neurônio morto" (dead neuron) que pode ocorrer com ReLU.
- Desvanecimento do gradiente: funções de ativação sigmoid e tanh: os gradientes durante o treinamento se tornam muito pequenos para atualizar os pesos da rede. Isso pode afetar o treinamento em redes profundas.
 - Neurônios mortos: não ativam para nenhum valor de entrada negativo. Isso pode acontecer quando um neurônio sempre retorna zero, impedindo que a rede aprenda.

Visão computacional: Tecnologias e projetos de visão computacional

- É usada em várias aplicações, como veículos autônomos, reconhecimento facial, análise de imagens médicas, realidade aumentada, automação industrial e muito mais. Além disso, a pesquisa continua a explorar novos domínios, como visão computacional multimodal.
- As CNNs revolucionaram a visão computacional. Projetos como AlexNet, VGG, Inception (GoogLeNet), ResNet e EfficientNet introduziram arquiteturas de redes profundas que superaram recordes em tarefas de reconhecimento de imagens, detecção de objetos e segmentação semântica.
- O OpenCV (Open Source Computer Vision Library) é uma das bibliotecas de visão computacional mais populares e muito utilizada.
 - O TensorFlow e PyTorch são duas das bibliotecas de aprendizado profundo mais usadas. Elas são usadas em uma variedade de projetos de visão computacional e oferecem uma estrutura flexível para construir e treinar modelos de CNN e outros tipos de redes neurais.

- O Yolo (You Only Look Once) é um projeto que introduziu uma abordagem eficiente para detecção de objetos em tempo real. Ele permite a detecção de objetos em imagens e vídeos com alta precisão e velocidade. O Mask R-CNN é um projeto que estende o Faster R-CNN para realizar também a segmentação semântica em imagens, é utilizado em aplicações médicas e de visão computacional em veículos autônomos.
- O ImageNet Large Scale Visual Recognition Challenge é um projeto que ofereceu uma grande base de dados rotulada com milhões de imagens e definiu benchmarks para tarefas de classificação de imagens.
 - Projetos de visão computacional são fundamentais em aplicações de robótica e Inteligência Artificial, incluindo veículos autônomos (como os da Waymo e Tesla), robôs industriais e assistentes de casa inteligente.

Planejamento de projeto de visão computacional

- Definição de objetivos: os objetivos podem implicar em tarefas como detecção de objetos, reconhecimento de padrões, segmentação de imagens, rastreamento de movimento, entre outros.
- Coleta de dados: imagens e vídeos capturados por câmeras, dados de sensores, imagens de satélite etc. para os objetivos do projeto.
- Análise de requisitos de hardware e software: poder de processamento, memória e capacidade de armazenamento, necessários para executar as tarefas de visão computacional. Escolha o software e as bibliotecas adequadas, como TensorFlow, OpenCV ou PyTorch, com base nas necessidades do projeto.
 - Design da arquitetura do sistema: a seleção de câmeras e sensores, configuração de hardware, escolha de algoritmos de visão computacional e estrutura de software. Certifique-se de que o sistema seja escalável e eficiente.

- Anotação de dados e treinamento de modelos: se o projeto envolver aprendizado de máquina, será necessário anotar os dados de treinamento, ou seja, rotular manualmente os objetos de interesse nas imagens. Em seguida, treine os modelos de visão computacional usando esses dados anotados.
- Desenvolvimento de software: isso pode incluir a implementação de algoritmos, a criação de interfaces de usuário, a integração com outros sistemas e a criação de APIs para acessar os resultados.
- Testes e validação: isso inclui testes de precisão, robustez, escalabilidade e tempo de resposta.
 - Integração e implantação.
 - Manutenção e atualizações.
 - Considerações éticas e legais: esteja ciente das considerações éticas e legais relacionadas ao projeto, especialmente se envolver captura de dados pessoais, reconhecimento facial ou questões de privacidade.

Projeto (continuação)

- Orçamento e prazos.
 - Documentação e treinamento.
 - Monitoramento e melhoria contínua.
-
- A organização das etapas de desenvolvimento de um projeto de visão computacional é fundamental para garantir que o projeto seja bem-sucedido, eficiente e cumpra seus objetivos.

Organização das etapas de projeto

- Definição do problema: a definição do problema é o ponto de partida fundamental. Nessa etapa, você precisa não apenas identificar o problema, mas também entender o contexto, os requisitos e as metas do projeto.
- Aquisição de dados: a qualidade e a quantidade dos dados adquiridos desempenham um papel crítico no sucesso do projeto.
- Pré-processamento de dados: no pré-processamento, os dados brutos são preparados para análise. Isso inclui ajustes de contraste, equalização de histograma, remoção de ruído e normalização.
 - Extração de características: envolve a identificação e o isolamento de informações relevantes de uma imagem ou vídeo, que podem ser usadas para descrever e distinguir objetos, padrões ou regiões de interesse.

Organização das etapas de projeto

As técnicas comuns de extração de características usando processamento de imagem:

- Detecção de borda: a detecção de borda envolve a identificação de transições bruscas de intensidade de pixel em uma imagem. Isso é feito por meio da aplicação de operadores, como o operador Sobel ou Canny, que realçam as bordas dos objetos na imagem. As bordas representam mudanças significativas nas características visuais e são frequentemente usadas na segmentação de objetos.
- Extração de textura: a textura de uma imagem refere-se à variação da intensidade dos pixels em pequenas regiões da imagem. A extração de características de textura pode incluir o uso de matrizes de coocorrência, histogramas de níveis de cinza ou descritores estatísticos para representar a textura de uma região.
 - Detecção de pontos de interesse: pontos de interesse, como cantos e keypoints, são locais distintivos na imagem que podem ser usados para rastreamento, alinhamento de imagens e correspondência. Algoritmos como o Harris Corner Detector e o SIFT (Scale-Invariant Feature Transform) são usados para identificar esses pontos.

Organização das etapas de projeto: Etapa de extração de características

- Histogramas de cores: os histogramas de cores representam a distribuição das cores na imagem.
- Momentos de imagem: os momentos de imagem são descritores estatísticos que resumem a distribuição de intensidades de pixel na imagem.
- Transformações geométricas e morfológicas: técnicas de transformação geométrica, como transformações de afinidade e rotações, podem ser usadas para extrair características relacionadas à geometria dos objetos na imagem. Além disso, operações morfológicas, como erosão e dilatação, podem destacar características estruturais.
- Filtros de resposta de frequência: como o filtro de Gabor.
 - Segmentação de regiões de interesse: a segmentação divide a imagem em regiões ou objetos distintos. As características extraídas de regiões segmentadas podem incluir área, perímetro, centroide e outros descritores relacionados à forma.

Organização das etapas de projeto

Seleção de algoritmos:

- Treinamento de modelos (se aplicável): se o projeto envolver aprendizado de máquina, como redes neurais, é necessário treinar modelos com dados etiquetados.
- Implementação do software: a implementação do software envolve a integração dos algoritmos e modelos em uma aplicação funcional.
- Avaliação e testes: a etapa de avaliação e testes é crucial para garantir que o sistema atenda aos requisitos. Isso inclui a realização de testes rigorosos, medição de desempenho, validação cruzada e a comparação dos resultados com métricas predefinidas. O feedback dos testes é usado para ajustar o sistema e melhorar seu desempenho.
 - Otimização e ajustes: com base nos resultados dos testes, o sistema é otimizado e ajustado. Isso pode incluir otimizações de código, afinamento de hiperparâmetros de algoritmos e refinamento de técnicas de pré-processamento.

Aplicações de reconhecimento facial

- Segurança e vigilância: o uso mais comum do reconhecimento facial é em sistemas de segurança, como controle de acesso a edifícios, aeroportos e eventos. Ele pode identificar pessoas autorizadas e detectar criminosos procurados.
- Autenticação biométrica: o reconhecimento facial é utilizado para autenticação em dispositivos móveis e sistemas de segurança, substituindo senhas tradicionais.
- Monitoramento de multidões: em eventos esportivos, concertos e manifestações públicas, o reconhecimento facial pode ser usado para monitorar multidões em tempo real, ajudando na segurança e na identificação de pessoas desaparecidas.
 - Marketing personalizado: identificar o perfil demográfico de seus clientes em lojas e shoppings, permitindo direcionar anúncios personalizados.
 - Assistência médica: em hospitais, o reconhecimento facial pode ajudar a identificar pacientes e acessar seus registros médicos com segurança, melhorando o atendimento ao paciente.

Aplicações de segmentação de objetos

- Detecção de veículos autônomos: a segmentação de objetos é crucial em veículos autônomos para identificar pedestres, outros veículos e obstáculos na estrada.
- Saúde e medicina: na medicina, a segmentação de objetos é usada para identificar e rastrear órgãos e estruturas do corpo em imagens médicas, como ressonâncias magnéticas e tomografias.
- Agricultura de precisão: agricultores usam a segmentação de objetos para identificar pragas, doenças e o crescimento das culturas, permitindo intervenções mais eficientes.
- Manufatura inteligente: a segmentação de objetos é usada em sistemas de inspeção de qualidade para identificar defeitos em produtos e embalagens de forma rápida e precisa.
 - Realidade aumentada: a segmentação de objetos é usada para sobrepor objetos virtuais ao mundo real de forma precisa.
 - Varejo e logística: no varejo, a segmentação de objetos é usada para rastrear produtos em estoque, automatizar processos de gerenciamento de inventário e facilitar a automação na logística.

Arcabouço tecnológico

- OpenCV (Open Source Computer Vision Library): o OpenCV é uma plataforma de código aberto utilizada para visão computacional. Ele oferece uma ampla variedade de algoritmos e funções para processamento de imagens, detecção de objetos, rastreamento, calibração de câmera e muito mais. O OpenCV é compatível com várias linguagens de programação, incluindo Python, C++, Java e outros.
- TensorFlow: desenvolvido pela Google, o TensorFlow é uma plataforma de aprendizado de máquina que também oferece suporte para visão computacional. TensorFlow Object Detection API.
 - PyTorch: PyTorch popularidade. Ele oferece suporte sólido para visão computacional com módulos como torchvision, que inclui conjuntos de dados, modelos pré-treinados ferramentas para treinar redes neurais para tarefas de visão.
 - Microsoft Azure Computer Vision: a plataforma Azure da Microsoft fornece serviços de visão computacional na nuvem. O Azure Computer Vision API permite que os desenvolvedores integrem recursos de reconhecimento de imagem em suas aplicações com facilidade.

Arcabouço tecnológico: Bibliotecas

- Dlib: é uma biblioteca C++ com várias funcionalidades, incluindo detecção facial, detecção de pontos de referência faciais, rastreamento de objetos e muito mais. É utilizado em aplicações de visão computacional e aprendizado profundo.
- Scikit-image: essa biblioteca Python é baseada no scikit-learn e fornece uma coleção de algoritmos para processamento de imagem, como filtragem, segmentação, transformações e muito mais.
- SimpleCV: é uma biblioteca Python que simplifica o desenvolvimento de aplicativos de visão computacional. Ela inclui uma variedade de ferramentas para trabalhar com imagens e vídeos, tornando mais fácil a implementação de tarefas comuns.

Arcabouço tecnológico: Linguagens de programação

- Python: é a linguagem de programação mais usada na visão computacional devido à sua facilidade de uso, vasta quantidade de bibliotecas de suporte e comunidade ativa de desenvolvedores. Bibliotecas como OpenCV, scikit-image e TensorFlow têm interfaces Python.
- C++: é frequentemente usado quando é necessária alta performance, como em sistemas embarcados e aplicações em tempo real. O OpenCV, por exemplo, oferece suporte nativo para C++.
- Java: é usado principalmente em aplicativos Android que envolvem visão computacional. A OpenCV oferece uma API Java para desenvolvimento em dispositivos Android.
 - C#: a Microsoft oferece suporte ao desenvolvimento de aplicativos de visão computacional em C# por meio das bibliotecas do Azure Computer Vision e da Emgu.CV, uma interface C# para o OpenCV.

Interatividade

A extração de características é um componente fundamental em projetos de visão computacional. Envolve a identificação e o isolamento de informações relevantes de uma imagem ou vídeo, que podem ser usadas para descrever e distinguir objetos, padrões ou regiões de interesse. **Não** se trata de uma técnica de extração de características:

- a) Filtros de resposta de frequência.
- b) Pooling.
- c) Detecção de bordas.
- d) Extração de textura.
- e) Detecção de pontos de interesse.

Resposta

A extração de características é um componente fundamental em projetos de visão computacional. Envolve a identificação e o isolamento de informações relevantes de uma imagem ou vídeo, que podem ser usadas para descrever e distinguir objetos, padrões ou regiões de interesse. Não se trata de uma técnica de extração de características:

- a) Filtros de resposta de frequência.
- b) **Pooling.**
- c) Detecção de bordas.
- d) Extração de textura.
- e) Detecção de pontos de interesse.

Referências

- LESSA, V. R. *Processamento de imagem e visão computacional*. Livro-Texto. Editora Sol: São Paulo, 2023.
- SOLOMON, C.; BRECKSON, T. *Fundamentos de processamento digital de imagens: uma abordagem prática com exemplos em Matlab*. 1. ed. Rio de Janeiro: LTC, 2013.
- UMBAUGH, S. E. *Digital image processing and analysis: human and computer vision applications with CVPTools*. Florida: CRC Press, 2011.

ATÉ A PRÓXIMA!