

Classificação de Imagens, Características de Histogramas e aprimoramento das imagens no domínio espacial com o dataset CIFAR-10

Narciso de Sousa Rodrigues¹; Pedro Azevedo A. de Oliveira¹; Saul Sousa da Rocha¹;

1. INTRODUÇÃO

A base de imagens escolhida foi a CIFAR-10, um *dataset* dividido em cinco *batches* de treinamento e um *batch* de teste, cada um com 10.000 imagens. O *batch* de teste contém exatamente 1.000 imagens selecionadas aleatoriamente de cada classe. Os *batches* de treinamento contém as imagens restantes em ordem aleatória, mas alguns *batches* de treinamento podem conter mais imagens de uma classe do que de outra. Entre eles, os *batches* de treinamento contém exatamente 5.000 imagens de cada classe. As classes encontradas no *dataset* são: avião, automóvel (carros pequenos), pássaro, gato, cachorro, veado, sapo, cavalo, navio e caminhão.

2. METODOLOGIA

Visando a aquisição de conhecimentos a cerca o tema visão computacional, as metodologias solicitadas para este trabalho foram: aquisição de histograma de cada imagem, a partir destes histogramas, foram obtidas as características de cada um destes histogramas: média(equação 1), variância(equação 2), skewness(assimetria), kurtosis(curtose), energia e entropia.

2.1 Histograma

Um histograma é uma espécie de gráfico de barras que demonstra uma distribuição de frequências. No histograma, a base de cada uma das barras representa uma classe e a altura representa a quantidade ou frequência absoluta com que o valor de cada classe ocorre. Ao mesmo tempo, ele pode ser utilizado como um indicador de dispersão de processos.

2.2 Média

A média (Me) é calculada somando-se todos os valores de um conjunto de dados e dividindo-se pelo número de elementos deste conjunto.

Como a média é uma medida sensível aos valores da amostra, é mais adequada para situações em que os dados são distribuídos mais ou menos de forma uniforme, ou seja, valores sem grandes discrepâncias.

$$Me = \frac{\sum(x * frq(x))}{n} \quad (1)$$

Onde:

Me: média

x: valor do pixel

frq(x): a quantidade de vezes que aparece o valor x

n: número de elementos do conjunto de dados

2.2 Variância

Dado um conjunto de dados, a variância é uma medida de dispersão que mostra o quão distante cada

valor desse conjunto está do valor central (médio). Quanto menor é a variância, mais próximos os valores estão da média; mas, quanto maior ela é, mais os valores estão distantes da média.

$$var = \sum((x - media)^2 * frq(x)) \quad (2)$$

Onde:

var: variância

x: valor do pixel

media: média do histograma

frq(x): a quantidade de vezes que aparece o valor x

2.3 Skewness(Assimetria)

A assimetria indica que os dados podem não ser normalmente distribuídos. Estes histogramas ilustram dados assimétricos. O histograma com dados assimétricos à direita mostra os tempos de espera. A maioria dos tempos de espera são relativamente curtos e apenas alguns tempos de espera são longos.

$$sk = \sum((x - media)^3 * frq(x)) / (\sum((x - media)^2 * frq(x)))^{\frac{3}{2}} \quad (3)$$

Onde:

sk: assimetria

x: valor do pixel

media: média do histograma

frq(x): a quantidade de vezes que aparece o valor x

2.4 Kurtosis(curtose)

Curtose significa o quanto de uma variável se encontra nas caudas da distribuição. A contribuição do pico ou do intervalo central para a curtose é pequena, embora essa ideia seja prevalente. Mas a curtose não mede a forma do pico – mas sim o “peso” das caudas, ou dos extremos da curva, sua fórmula se dá pela equação abaixo.

$$ku = \sum((x - media)^4 * frq(x)) / (\sum((x - media)^2 * frq(x))^2) \quad (4)$$

Onde:

ku: curtose

x: valor do pixel

media: média do histograma

frq(x): a quantidade de vezes que aparece o valor x

2.5 Energia

A energia de um histograma é uma medida da concentração dos valores do histograma. Ela é dada pela seguinte fórmula:

$$ene = \sum(frq(x)^2) \quad (5)$$

Onde:

ene: energia

x: valor do pixel

frq(x): a quantidade de vezes que aparece o valor x

2.6 Entropia

A entropia é uma medida da incerteza ou da informação contida em um sistema. No contexto de um histograma, a entropia pode ser calculada a partir da distribuição de frequências das diferentes classes ou bins do histograma.

$$ent = \Sigma(-prob * \log_2(prob))$$

Onde:

ent: Entropia

prob: probabilidade da classe(fração de ocorrências da classe sobre o número total de amostras)

Além dos atributos acima listados, as imagens para efeito de comparação foram modificadas com o aprimoramento das imagens no domínio espacial para buscar um resultado mais adequado do que o original. Para isso foi requisitado fazer o alargamento de contraste, o negativo, transformações de log e transformações da lei de potência. Diante de tais transformações pode-se perceber variações que serão melhor explicadas na etapa dos resultados.

2.7 Negativo

O negativo de uma imagem com níveis de cinza na faixa [0, L-1] é obtido usando a transformação negativa, dada pela função:

$$s = L - 1 - r$$

basicamente ela inverte os níveis de intensidade da imagem. Esse tipo de processamento é particularmente adequado para aprimorar detalhes em branco ou cinza embutidos em regiões escuras de uma imagem, especialmente quando as áreas pretas são dominantes em tamanho.

2.8 Transformações de log

Dada pela função:

$$s = c \log(1 + r)$$

Essa transformação mapeia uma faixa estreita de valores baixos de nível de cinza na imagem de entrada em uma faixa mais ampla de níveis de saída. Usamos uma transformação desse tipo para expandir os valores de pixels escuros em uma imagem enquanto comprimimos os valores de nível superior.

2.9 Transformações da Lei de Potência (Gamma)

Dada pela função:

$$s = cr^\gamma$$

onde c e γ são constantes positivas. A lei de potência com valores fracionários de γ mapeiam uma faixa estreita de valores de entrada escuros em uma faixa mais ampla de valores de saída. Basicamente a função irá suavizar as intensidades muito altas ou baixas da imagem.

2.10 Alargamento de Contraste

Uma das funções lineares por partes mais simples, dada pela função:

$$s = 1/(1 + (k/r)^e)$$

Imagens de baixo contraste podem resultar de iluminação ruim, falta de faixa dinâmica no sensor de imagem ou até mesmo configuração incorreta da abertura da lente durante a aquisição da imagem. A ideia por trás do alongamento de contraste é aumentar a faixa dinâmica dos níveis de cinza na imagem sendo processado.

3. RESULTADOS E DISCUSSÃO

Os histogramas resultantes mostram que as classes da nossa base de dados apresentam, em grande parte das imagens, uma frequência aproximada. E tendo poucas características diferenciais entre classes.

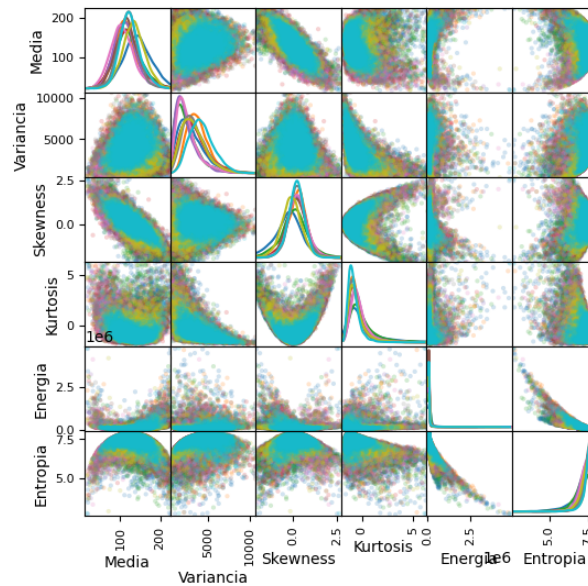


Imagem 1: Scatter de todos os histogramas atribuídos ao dataset CIFAR-10.

Dado os resultados presentes na tabela 1, percebemos que o modelo utilizado não foi capaz de gerar bons resultados na classificação da base de imagens CIFAR-10.

Função	Parâmetros	accuracy
Normal	-	27%
Negativo	-	10%
Logarítmico	-	38%
Alongamento de contraste	(k=1, e=16)	10%
Alongamento de contraste	(k=0.1, e=6)	10%
Alongamento de contraste	(k=1, e=2)	10%
Alongamento de contraste	(k=0.4, e=4)	10%
Potência	($\gamma = 3$)	10%
Potência	($\gamma = 4$)	10%
Potência	($\gamma = 5$)	10%
Potência	($\gamma = 0.4$)	10%

Tabela 1: Resultados das classificações.

Foi discutido que, um dos fatores que pode ter interferido no aprendizado do modelo está atribuído à resolução das imagens, pois o dataset é formado por imagens 32x32x3 que são consideradas pequenas, sendo

assim, a adição do aprimoramento das imagens no domínio espacial não surtiram tanto efeito.

Nosso melhor resultado foi com o melhoramento das imagens com a função logarítmica, alcançando uma precisão de 38%. A explicação para tal resultado é que a mesma deixou as imagens mais suavizadas.

4. CONCLUSÃO

Comparando as métricas propostas foi obtido às margens de 10% para 9 casos, 27% para o caso da execução na base de dados normal e 38% na base de dados com a transformação logarítmica, o que fez perceber que com adequação obteve-se baixa no desempenho de 28% mas teve ganho de 11%, deste modo abrindo possibilidades para melhorias de outras maneiras, que devem ser estudadas e aplicadas.

Como conclusão analisando o melhor caso em que tivemos 38% logo percebemos que a solução não obteve nem de perto uma margem satisfatória de resposta, deste modo ficando em aberto a necessidade de processamentos com métricas ou modelos diferentes para que se possa alcançar o mínimo de uma taxa pelo menos superior a 50%.

5. ANEXO

5.1 Tabela de classificação usando uma rede neural, PyTorch:

Função	Parâmetros	accuracy
Normal	-	64 %
Negativo	-	10 %
Logarítmico	-	64 %
Alargamento de contraste	(k=1, e=16)	37 %
Alargamento de contraste	(k=0.1, e=6)	42 %
Alargamento de contraste	(k=1, e=2)	64 %
Alargamento de contraste	(k=0.4, e=4)	63 %
Potência	($\gamma = 3$)	59 %
Potência	($\gamma = 4$)	57 %
Potência	($\gamma = 5$)	55 %
Potência	($\gamma = 0.4$)	62 %

Tabela 2: Resultados das classificações usando uma rede neural.