Identificação e Classificação de Imagens Publicitárias Quanto à Qualidade Utilizando Aprendizado de Máquina

Evandro Oliveira[1], **Ana Carolina Costa**[2], **Cícero Moura**[3]

[1]Universidade Fernando Pessoa - UFP - Porto - Portugal

[2]Universidade Federal do Ceará - UFC - Fortaleza - Ceará - Brasil

[3]Pontifícia Universidade Católica de Minas Gerais - PUC/MG - Belo Horizonte - Minas Gerais - Brasil

{evandro.oliveira, ana.costa, cicero.moura}@ionsistemas.com.br

Abstract. *This work suggests a prediction of the aesthetic quality of objective images with a method of classification of ruins according to the opinion of annotator. To this end, deep learning techniques are used in a set of synthetic advertising images. The quality assessment proposal is suitable for assessing the quality of existing methods on the quality scale. The results of the experiments show that ResNet50 is more efficient than the NIMA and Koncept512 models, reaching 33% improvement in relation to correlations and 11% improvement in relation to MicroF1.*

Resumo. *Este trabalho sugere um método para a predição da qualidade estética de imagens com o objetivo de filtrar imagens que seriam classificadas como ruins conforme a opinião do anotador. Para tal, são utilizadas técnicas de aprendizagem profunda em um conjunto de dados com imagens sintéticas de caráter publicitário. A abordagem sugerida é comparada com outros métodos já existentes para inferir a qualidade da imagem considerando o Coeficiente Pearson e Acurácia em conjuntos de imagens avaliadas numa escala de boa a ruim. Os resultados dos experimentos mostram que a ResNet50 se mostra mais eficiente do que os modelos NIMA e Koncept512, chegando em 33% de melhora em relação às correlações e melhora de 11% em relação ao MicroF1.*

1. Introdução

O crescimento de conteúdos visuais e textuais, em meio digital e criados por usuários, tornaram-se comuns. Este aumento impulsiona a necessidade de avaliar os conteúdos digitais a fim de aplicar filtros e remover materiais inadequados ao seu contexto. Entretanto, imagens sintéticas não são endereçadas em abordagens ou conjuntos de dados bem estabelecidos para inferência da qualidade.

Para isso, é comum encontrar plataformas de *software* com recursos de vídeos, fotos e textos criados pelos próprios usuários dessas aplicações. Em sistemas abertos ao público é desejável que tal conteúdo seja submetido a alguma análise antes da sua publicação em determinada plataforma. No contexto de imagens, existem trabalhos que buscam inferir a qualidade de uma imagem de forma automática, sem nenhuma referência, predizendo uma nota, que consiste em um valor escalar dentro de um intervalo fechado [Li et al. 2016, Esfandarani and Milanfar 2017, Hosu et al. 2019].

Dessa forma, os conjuntos de dados construídos para avaliação e treinamento de métodos que visam avaliar quantitativamente a qualidade de uma imagem contam com um grupo de anotadores que são pessoas que atribuem, cada um, um resultado para a imagem. No entanto, estas bases de dados não abrangem imagens sintéticas e direcionam as soluções a predizer a opinião média do grupo de anotadores sobre cada imagem.

Logo, são definidas como imagens sintéticas aquelas imagens produzidas via edição, unindo fotos, textos, etc. Este tipo de imagem está presente em todo meio digital, em especial no contexto publicitário. Tais imagens podem ser encontradas em *banners* em sites comerciais. Por esse motivo, tais imagens não são mostradas neste trabalho. Todavia, é de fundamental importância um método que seja capaz de avaliar imagens desta natureza e que não é tão explorada na literatura.

Dito isso, os experimentos realizados mostram que os principais métodos para inferência da qualidade da imagem, treinados em grandes conjuntos de dados já bem estabelecidos, falham em produzir bons resultados ao avaliar imagens sintéticas. Este comportamento foi evidenciado pela geração de notas pouco relacionadas com a anotação humana. Isso acontece porque o domínio das imagens destes conjuntos não inclui imagens sintéticas. Além disso, estas imagens possuem a anotação de diversos anotadores, gerando uma distribuição contínua da rotulagem a qual os métodos buscam aprender, ao contrário de situações práticas onde a rotulagem dos dados corresponde à classes bem definidas. Ademais as imagens não são redimensionadas para serem rotuladas.

Sendo assim, neste trabalho são sugeridos métodos para filtragem de imagens publicitárias em função da sua qualidade estética. O objetivo destes métodos é ser capaz de filtrar imagens que seriam classificadas como ruins conforme a opinião de um anotador. Para isso, foi reunido um grupo de 1493 imagens sintéticas de caráter publicitário, posteriormente rotuladas com uma nota numa escala de 0 a 4, onde 0 representa as imagens muito ruins e 4 representa as imagens muito boas. Os métodos apontados consistem em predizer esta nota de acordo com o viés do anotador e determinam um limiar para designar quais imagens são boas ou quais imagens são ruins.

Para atingir o objetivo principal, na Seção 2 é abordado o referencial teórico que foi levantado para esta pesquisa. Na Seção 3 é mostrada a metodologia com a arquitetura geral dos modelos a serem experimentados, apresentando detalhadamente como foi feita essa coleta. A partir da metodologia, na Seção 4 são mostrados os resultados acerca dos experimentos realizados com a classificação de imagens e na Seção 5 são feitas as considerações finais a respeito do uso das técnicas apresentadas para classificação das imagens.

2. Trabalhos relacionados

Dada a sua importância, a classificação de dados digitais em relação ao seu conteúdo e/ou qualidade é uma tarefa bem estudada [Wang et al. 2019, Shahid et al. 2014, Ortis et al. 2019, Liu and Forss 2015]. Se tratando de conteúdos textuais, alguns trabalhos buscam identificar aqueles que seriam inadequados a determinados contextos [Pelle et al. 2018, Pitsilis et al. 2018, Schmidt and Wiegand 2017], enquanto outras abordagens buscam realizar uma análise de sentimento baseada no teor da mensagem [Xu et al. 2019, Mohey El-Din 2016]. Embora ambas as tarefas possam ser utilizadas para identificar e filtrar conteúdos de forma específica, o processamento de linguagem

natural, com esse objetivo, pode ser auxiliado pela identificação de palavras chaves, que direcionam fortemente o teor de um texto [Papagiannopoulou and Tsoumakas 2020].

Acerca de conteúdos visuais, a inferência da qualidade da imagem é também uma tarefa bem estabelecida [Zhai and Min 2020]. As bases de dados que endereçam essa realização contam com a anotação de diversas pessoas [Murray et al. 2012, Zhang et al. 2018, Isola et al. 2017, Hosu et al. 2019], gerando assim uma tendência geral sobre uma imagem. Consequentemente, os trabalhos existentes focam em predizer a opinião média sobre uma imagem e não se preocupam em capturar especificidades na opinião de um único anotador. Além disso, estes conjuntos dados possuem poucas imagens sintéticas, focando em imagens realistas, onde os objetos centrais são pessoas, animais ou paisagens. Dessa forma, os conjuntos de dados existentes para a tarefa de inferência da qualidade da imagem não contemplam imagens sintéticas, tampouco imagens publicitárias, e são voltados para cenários onde o interesse está em predizer a opinião média do conjunto de anotadores.

Visando predizer uma nota para a qualidade subjetiva de uma imagem, alguns autores utilizam redes neurais convolucionais para a construção de métodos que sejam capazes de produzir bons resultados dados existentes. Talebi e Milanfar propõem uma rede neural, NIMA [Esfandarani and Milanfar 2017], capaz de predizer uma pontuação referente a opinião média acerca de uma imagem e uma curva de distribuição dessa pontuação, baseada no conjunto de anotadores de tais dados em questão.

Entretanto, no cenário de interesse deste trabalho não é possível determinar uma curva de distribuição para o treinamento do modelo proposto. Semelhante a esta abordagem, Hose *et al.* propõem uma rede neural convolucional, chamada de Koncept512 [Hosu et al. 2019], que utiliza como *backbone* a rede InceptionResNetV2 [Szegedy et al. 2016] para, assim como o modelo NIMA, predizer uma pontuação de opinião média e uma distribuição dessa pontuação. Este método, além de não endereçar casos onde não existe uma distribuição das anotações para o treinamento, configura uma solução muito custosa computacionalmente.

Apesar dos resultados demonstrados nos trabalhos citados, os métodos propostos não parecem se mostrar capazes de performar um bom balanço entre qualidade e tempo de inferência. Neste trabalho é priorizada uma abordagem simples que seja capaz de produzir resultados relevantes e gerar respostas com certa agilidade.

3. Metodologia

Para a avaliação e o treinamento de métodos capazes de englobar a inferência da qualidade de imagens sintéticas, inicialmente foi elaborado um conjunto de 1493 imagens associadas a uma nota entre 0 e 4. Essa rotulação é feita através dos três anotadores que elaboraram o *dataset*, no qual estes anotadores são profissionais de Tecnologia da Informação (TI) com experiência em imagens publicitárias na área de Logística.

Tais imagens foram avaliadas pelos três anotadores e o resultado segue a avaliação da maioria, por exemplo: se os três anotadores avaliarem que uma imagem tem nota 3, ela será 3. Porém se todos os anotadores colocarem uma nota diferente, a nota será a do meio (exemplo: se o anotador A der nota 1, o anotador B der nota 2 e o anotador C der nota 3, prevalecerá a nota 2 para a imagem). A nota por voto da maioria também é considerado.

Sendo assim, a Figura 1 mostra a distribuição das classes do conjunto reunido. Neste gráfico de barras percebe-se que é notória a maior concentração de imagens com notas medianas e a ausência de imagens ou muito boas ou muito ruins, confirmando o padrão condizente com as situações reais.

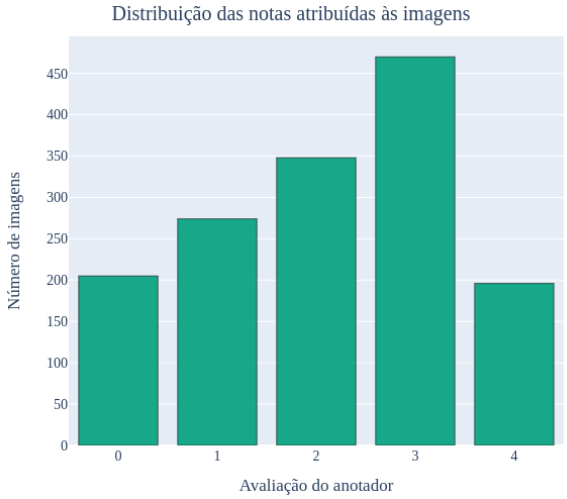

Figura 1. Distribuição das classes

Ademais, as imagens selecionadas possuem resoluções variadas como visto na Figura 2. Para o treinamento dos modelos sugeridos por este trabalho é utilizada uma resolução de 480x192 *pixels*, compatível com o *aspect ratio* (razão entre largura e altura da imagem) médio do conjunto.

A base de imagens construída é focada em figuras sintéticas e publicitárias. Dado este cenário, durante a elaboração do conjunto foi necessário garantir a seleção de imagens de uma mesma temática assim como variabilidade na anotação, já que sem isso, para os métodos de aprendizado, seria possível criar um grande viés no resultado. Deste ponto do texto em diante estas imagens serão referenciadas como **C**onjunto de **I**magens **S**intéticas **P**ublicitárias, ou **CISP**.

Dessa forma, na Figura 1, as imagens selecionadas para compor o CISP possuem avaliações bem distribuídas no intervalo definido. As classes nos extremos (0 e 4) possuem um número menor de amostras visto a menor frequência de imagens qualificadas como muito ruins ou muito boas.

Já na Figura 2 é apresentada as características dos dados no CISP e, embora as imagens selecionadas contenham aspectos variados, é notável a predominância de imagens com *aspect ratio* maior que 1, isto é, imagens com largura maior que a altura.

Como as anotações podem ser aglutinadas em classes bem separadas, a primeira abordagem experimentada consiste em utilizar a Resnet50 [He et al. 2016] pré-treinada no ImageNet [Zhu et al. 2017] como *backbone* para uma rede neural para classificação entre 5 classes. Junto a essa abordagem, foi testado também o mesmo *backbone* porém realizando a classificação entre duas classes: imagens boas (nota maior ou igual a 2) e imagens ruins (nota menor que 2). A arquitetura desses modelos pode ser vista na Figura 3. Ambos os modelos são treinados com a função de perda *Categorical Cross Entropy*. Finalmente, foi experimentada mesma arquitetura, agora possuindo uma única

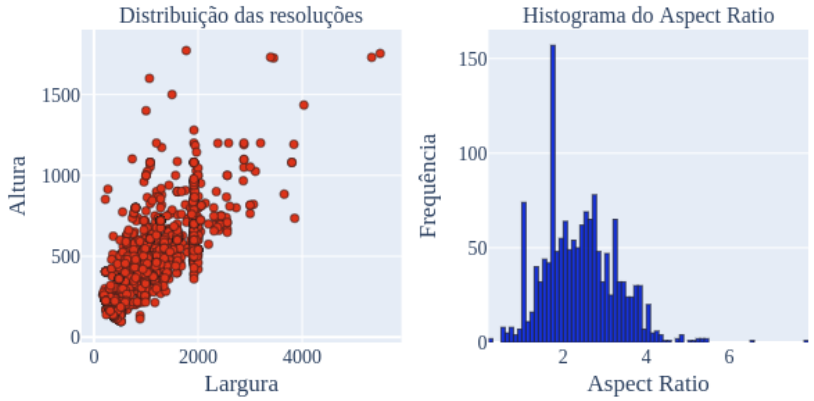

Figura 2. Características dos dados no CISP

saída, correspondente a nota atribuída à imagem. Este modelo, por não se tratar de uma tarefa de classificação, foi treinado utilizando a função de perda *Mean Squared Error*.

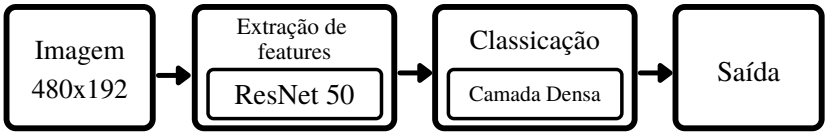

Figura 3. Arquitetura geral dos modelos experimentados

Na Figura 3 foi priorizada uma arquitetura simples e ágil, permitindo adaptação e retreinamento conforme a necessidade. A camada densa do modelo foi alterada em cada experimento para produzir o número de saídas de acordo com a abordagem.

Todos os modelos foram treinados por 10 épocas, utilizando *batchs* de 16 imagens redimensionadas para 480x192 *pixels*. Também, foi utilizado o optimizador *Adam* com uma taxa inicial de aprendizado de $3,5 \times 10^{-4}$ reduzida em um fator de 0,7 a cada 10 épocas. A fim de evitar *overfitting*, que é quando o modelo praticamente decora os dados, performando bem nos dados de teste, mas cai em relação aos dados de treino, é requerido aumentar a capacidade de generalização dos modelos. Para isso, foram utilizadas técnicas de *data augmentation*, que é a técnica de aumentar a quantidade e diversificar os dados, durante o treinamento. O objetivo dessa técnica é equiparar as quantidades de cada classe a fim de que o modelo treine igualmente para todas as notas atribuídas. Portanto, ficou cada classe com 450 amostras, como mostrado na tabela a seguir:

Tabela 1. Quantidade de cada classe da base de dados.

classe	dados sintéticos	data augmentation	total
0	203	247	450
1	274	176	450
2	349	101	450
3	474	-24	450
4	196	254	450

Por outro lado, pela natureza do problema, não é possível aproveitar de técnicas que degradam a qualidade da imagem (ruído, distorção, borrões, etc.) visto que poderiam atrapalhar o treinamento inserindo informações contradizentes. Desta forma, as técnicas de *data augmentation* selecionadas foram a Rotação Aleatória, de no máximo de 10°, e a Inversão Horizontal Aleatória, com probabilidade de 20%.

Por fim, o CISP foi dividido em uma partição de treino com 70% dos dados, uma partição de validação e uma partição de teste, cada uma correspondendo a 15% da base, mantendo a proporção entre classes em todas as partições. Os modelos selecionados durante o treinamento correspondem a menor perda na partição de validação.

4. Resultados

Nesta seção são apresentados os resultados das abordagens discutidas. Para tanto, são comparadas as performances dos modelos utilizando a ResNet50 e dos métodos existentes na literatura. Estes últimos são treinados em conjuntos de dados já estabelecidos chamados de AVA e KonIQ, assim como no CISP. Os treinamentos efetuados na base CISP seguem a metodologia descrita na Seção 3. Além disso, é feita uma comparação dos tempos de inferência dos métodos, a fim de analisar o balanço entre a performance e o custo.

Durante a avaliação das abordagens, é analisada a capacidade dos modelos de produzir respostas condizentes com o viés do anotador. A Tabela 2 apresenta os resultados dos métodos citados e sugeridos obtidos na partição de teste do CISP. São reportadas as correlações de Pearson (PLCC) e de Spearman (SRCC) para fins comparativos, além das métricas de Acurácia e de Micro F1. Estas últimas são referentes à capacidade de divisão do CISP em dois conjuntos: imagens classificadas como boas (nota maior ou igual 2) e imagens classificadas como ruins (nota menor que 2). Portanto, nos modelos onde ocorre a predição de uma nota é definido um limiar para divisão das amostras otimizado na partição de validação.

Tabela 2. Performance dos métodos citados e sugeridos obtidos na partição de teste do CISP.

Método	Métricas			
	PLCC	**SRCC**	**Acurácia**	**Micro F1**
ResNet50 - 5 classes	0.62	0.62	59.5%	**0.81**
ResNet50 - 2 classes	-	-	59.8%	**0.81**
ResNet50 - 1 classe	**0.68**	**0.68**	61.6%	**0.81**
NIMA (AVA)	0.37	0.35	56.2%	0.71
NIMA (CISP)	0.59	0.59	50.0%	0.75
Koncept512 (KonIQ)	0.30	0.29	**65.6%**	0.70
Koncept512 (CISP)	0.62	0.61	53.1%	0.75

Na Tabela 2 é evidenciado a ineficiência de generalização das bases AVA e KonIQ para o contexto proposto neste trabalho, atestado por coeficientes de correlação baixos. Apesar disso, é perceptível que a habilidade destes em dividir a partição de teste do CISP em dois conjuntos distintos não é baixa quando comparada com os modelos treinados no próprio CISP. Estes modelos apresentam um ganho quando treinados na partição de

treino do CISP, o que atesta o fato das bases já existentes não contemplarem o cenário de imagens sintéticas.

Embora tenha sido demonstrada uma melhora quando alterada a partição de treino, os modelos NIMA e Koncept512 não representam os melhores resultados. Por se tratarem de redes neurais mais complexas que os outros modelos avaliados, os valores reportados podem ser explicados pela baixa quantidade de dados do conjunto de treino. Entretanto, a baixa quantidade de amostras é comparável com situações reais, visto que o custo de anotação e seleção de imagens para compor uma base é alto. Deste modo, estas abordagens não se apresentaram adequadas para o cenário proposto pelo trabalho.

Os métodos sugeridos por este trabalho, utilizando a ResNet50 e alterando somente a camada de classificação e a função de perda, obtiveram resultados tão bons quanto, e até melhores, que os modelos NIMA e Koncept512. O modelo que utiliza a ResNet50 com uma saída de 5 classes performa tão bem quanto o Koncept512 e superior ao NIMA, em relação aos coeficientes de correlação. Já a ResNet50 com somente 1 saída e treinada utilizando *Mean Absolute Error* apresenta os melhores coeficientes de correlação, indicando que embora o conjunto de dado contenha as anotações bem discretizadas (Figura 4), o modelo se beneficia da liberdade em posicionar as predições em um espectro contínuo.

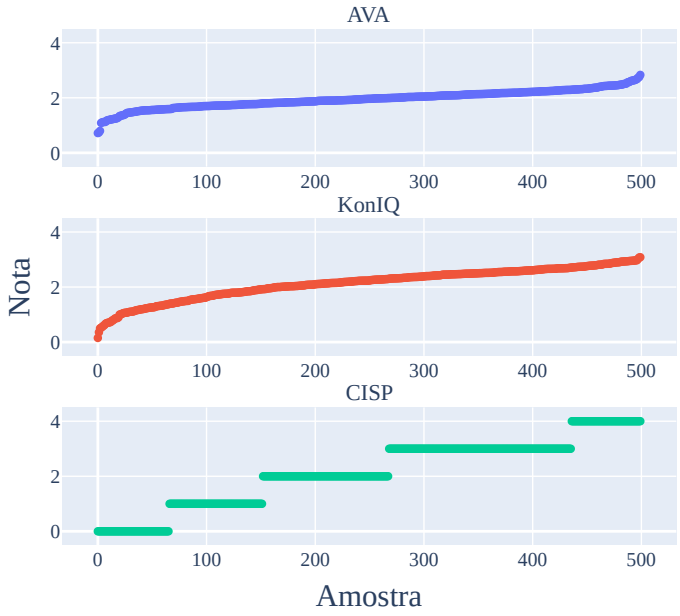

Figura 4. Comparação entre as predições alvo dos conjuntos de imagens

Observando as métricas associadas a capacidade de divisão do conjunto de teste em dois, é possível perceber que o modelo utilizado para classificação binária, ResNet50 com 2 classes de saída, apesar de ser treinado especificamente para esta tarefa, não apresenta o melhor resultado. Este comportamento pode ser explicado pela natureza dos dados, já que as anotações não fornecem informação suficiente para que o modelo consiga aprender, utilizando uma função de perda de classificação e relações entre imagens de uma mesma classe, como pode ser observado na Figura 4, que mostra a seleção de 500 amostras aleatórias de cada conjunto e realizada a ordenação crescente de suas notas.

Dessa forma, a capacidade de generalização do modelo é prejudicada, pois nesta hipótese é reforçada pelos resultados apresentados por aqueles modelos que realizam a predição de uma nota para cada imagem. Logo, por serem treinados utilizando funções de perda de regressão, os modelos possuem liberdade para posicionar as imagens em um espectro de notas, não se limitando a somente duas opções.

No cenário levantado por este trabalho, onde o objetivo consiste em determinar uma abordagem capaz de performar predições consistentes com o viés de um anotador, além da capacidade do modelo em produzir respostas adequadas, é necessário avaliar o custo computacional.

Para tal, a Tabela 3 provê uma comparação entre os tempos de inferência dos modelos. Foram realizadas 500 medições com cada modelo, sempre utilizando as mesmas amostras e resolução de entrada dos modelos de 480×192 . O resultado reportado consiste da média e desvio padrão. Todas as medições foram realizadas em *CPU*, em um processador Intel Core i5-10210U (4x 1.6GHz) e 16GB DDR4 RAM.

Tabela 3. Comparação entre os tempos de inferência dos modelos.

Método	Tempo (ms)
ResNet50 - 5 classes	124 ± 14
ResNet50 - 2 classes	136 ± 20
ResNet50 - 1 classe	130 ± 12
NIMA	305 ± 50
Koncept512	232 ± 41

Em cenários reais de aplicação da tarefa, a latência dos modelos pode ser determinante para a escolha da abordagem a ser utilizada. De acordo com os resultados apresentados Tabela 3 os modelos NIMA e Koncept512 apresentam maior tempo de inferência, não justificado pelos resultados apresentados no CISP, mesmo quando treinados no mesmo. A inferência também poderia ter sido realizada na GPU, no qual o tempo provavelmente se tornaria irrelavante.

No entanto, no cenário atual, os métodos que utilizam a ResNet50 realizam a inferência em tempos bem próximos. Sendo assim, o modelo mais indicado consiste na ResNet50 com 1 classe de saída, já que apresenta o melhor balanço entre performance e agilidade.

5. Conclusão

Tendo em vista os experimentos e resultados mencionados, vê-se que é realizada uma avaliação dos métodos e das bases de dados voltados para a tarefa de inferência da qualidade da imagem em um cenário onde o interesse consiste em avaliar imagens sintéticas publicitárias. Após identificar que os conjuntos de imagens existentes não contemplam dados similares aos do contexto proposto, foi construído o Conjunto de Imagens Sintéticas Publicitárias, o CISP, para treinamento e avaliação de métodos em uma base adequada. Este conjunto possui 1493 imagens, divididas entre 5 classes de anotação, correspondentes a nota designada por um avaliador.

Os experimentos realizados visam avaliar a capacidade dos métodos em produzir

respostas altamente correlacionadas com o viés de um anotador além de ser capaz de separar as amostras em conjuntos distintos, sendo um deles correspondente a amostras boas e ruins. Uma das aplicações da tarefa de inferência da qualidade da imagem consiste em filtrar dados baseado na avaliação predita pela abordagem. Portanto, é essencial avaliar a capacidade dos modelos em performar tal tarefa.

Desse modo, os resultados obtidos reforçam o fato das bases de dados já existentes não contemplarem o cenário de imagens sintéticas publicitárias, visto que as métricas de correlação obtida nestes casos são inferiores aos modelos treinados no CISP. Portanto, utilizando a ResNet50, as métricas de correlação calculadas representam resultados melhores ou iguais aos obtidos com os modelos NIMA e Koncept512, podendo chegar a 33% melhor com os experimentos realizados, o que pode ser explicado pela baixa quantidade de dados da partição de treino, não contendo informação suficiente para o aprendizado destes últimos. Contudo, em função do custo de se gerar anotações de uma base de dados em um cenário real, a partição de treino do CISP corresponde a um cenário realista.

Finalmente, foi observado um desempenho superior no método que utiliza a ResNet50 e uma única classe de saída treinado utilizando uma função de perda voltada para regressão. Este resultado indica que o modelo se beneficia da liberdade em sua saída, não sendo necessário selecionar uma entre as cinco classes para avaliar uma nova amostra. Além disso, tendo como base a avaliação de tempo de inferência, o modelo que utiliza a ResNet50 e 1 classe de saída demonstra um ótimo balanço entre performance e agilidade, sendo, no cenário proposto, o método mais indicado.

Referências

Esfandarani, H. T. and Milanfar, P. (2017). NIMA: neural image assessment. *CoRR*, abs/1709.05424.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778.

Hosu, V., Lin, H., Sziranyi, T., and Saupe, D. (2019). Koniq-10k: An ecologically valid database for deep learning of blind image quality assessment. *CoRR*, abs/1910.06180.

Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134.

Li, Y., Po, L., Feng, L., and Yuan, F. (2016). No-reference image quality assessment with deep convolutional neural networks. pages 685–689.

Liu, S. and Forss, T. (2015). New classification models for detecting hate and violence web content. In *2015 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K)*, volume 01, pages 487–495.

Mohey El-Din, D. (2016). A survey on sentiment analysis challenges. *Journal of King Saud University - Engineering Sciences*, pages –.

Murray, N., Marchesotti, L., and Perronnin, F. (2012). Ava: A large-scale database for aesthetic visual analysis. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2408–2415.

Ortis, A., Farinella, G., and Battiato, S. (2019). An overview on image sentiment analysis: Methods, datasets and current challenges. pages 290–300.

Papagiannopoulou, E. and Tsoumakas, G. (2020). A review of keyphrase extraction. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(2):e1339.

Pelle, R., Alcântara, C., and Moreira, V. (2018). A classifier ensemble for offensive text detection. pages 237–243.

Pitsilis, G. K., Ramampiaro, H., and Langseth, H. (2018). Detecting offensive language in tweets using deep learning. *CoRR*, abs/1801.04433.

Schmidt, A. and Wiegand, M. (2017). A survey on hate speech detection using natural language processing. In *Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media*, pages 1–10, Valencia, Spain. Association for Computational Linguistics.

Shahid, M., Rossholm, A., Lövström, B., and Zepernick, H.-J. (2014). No-reference image and video quality assessment: a classification and review of recent approaches. *EURASIP Journal on Image and Video Processing*, 2014.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2818–2826.

Wang, B., Chen, B., Ma, L., and Zhou, G. (2019). User-personalized review rating prediction method based on review text content and user-item rating matrix. *Information*, 10(1).

Xu, G., Yu, Z., Yao, H., Li, F., Meng, Y., and Wu, X. (2019). Chinese text sentiment analysis based on extended sentiment dictionary. *IEEE Access*, 7:43749–43762.

Zhai, G. and Min, X. (2020). Perceptual image quality assessment: a survey. *Science China Information Sciences*, 63:1–52.

Zhang, R., Isola, P., Efros, A. A., Shechtman, E., and Wang, O. (2018). The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595.

Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232.