

Classificação de Escudos da Premier League em Ambientes Reais com Redes Neurais Convolucionais*

1st Sergio A. Cezar, 2nd Yandre M. G. Costa

*Departamento de Informática, Universidade Estadual de Maringá
ra134680@uem.br, yandre@din.uem.br*

Abstract—Reconhecer logotipos esportivos em ambientes reais é uma tarefa desafiadora devido à variabilidade de textura, iluminação, deformação e oclusão. Neste trabalho, desenvolveu-se um classificador para os 20 clubes da Premier League utilizando uma abordagem de Transfer Learning com Fine-Tuning sobre a ResNet18 pré-treinada no ImageNet. Um dataset híbrido foi construído com 150 imagens sintéticas (escudos em fundo branco) e 150 imagens reais por classe, contendo camisas, bonés, faixas, acessórios e objetos diversos. Para evitar overfitting no domínio sintético, aplicou-se uma estratégia agressiva de Data Augmentation utilizando a biblioteca Albumentations, com transformações geométricas, cromáticas e de oclusão. Após 15 épocas de treinamento, o modelo atingiu 88% de acurácia no conjunto de validação, indicando boa capacidade de generalização para condições reais.

Index Terms—Transfer Learning, Data Augmentation, ResNet18, Classificação de Imagens.

I. INTRODUÇÃO

O reconhecimento de logotipos é uma subárea importante da visão computacional, com aplicações que incluem marketing esportivo, análise de transmissão televisionada, monitoramento de torcida, reconhecimento de times em redes sociais e sistemas automáticos de catalogação. Ao contrário de objetos naturais, logotipos possuem formas bem definidas, porém sofrem drásticas deformações no mundo real quando aplicados em tecidos, acessórios e superfícies não planas.

Modelos treinados apenas com imagens limpas, por exemplo, escudos em fundo branco, apresentam dependência do contexto visual e tendem a falhar em ambientes reais. Assim, o desafio principal deste trabalho é a redução do domain gap entre imagens sintéticas e reais.

Este artigo apresenta uma abordagem completa que integra (i) construção de dataset híbrido, (ii) Transfer Learning com Fine-Tuning total, e (iii) Data Augmentation avançado utilizando Albumentations.

II. FUNDAMENTAÇÃO TEÓRICA

Para compreender as soluções adotadas neste trabalho, é necessário apresentar os conceitos fundamentais relacionados ao uso de redes neurais convolucionais e técnicas modernas de aprendizado profundo.

A. Redes Neurais Convolucionais (CNNs)

As Redes Neurais Convolucionais são amplamente utilizadas em tarefas de classificação de imagens. Diferentemente de redes totalmente conectadas, elas preservam a estrutura espacial da imagem ao utilizar filtros (kernels) que percorrem a imagem em pequenas regiões, extraíndo padrões visuais relevantes.

Hierarquia de Características: as CNNs aprendem características de forma hierárquica:

- Camadas iniciais: identificam bordas, linhas e cores básicas.
- Camadas intermediárias: detectam formas e texturas mais complexas.
- Camadas finais: reconhecem partes específicas de objetos, permitindo a classificação completa, como diferentes escudos de times.

B. Arquitetura ResNet18

A ResNet18 foi utilizada neste trabalho por ser leve e eficiente. Ela utiliza conexões residuais (skip connections), que ajudam a treinar redes profundas evitando a perda de gradiente. Essa característica torna a ResNet adequada para problemas com poucos dados, como é o caso deste estudo.

C. Transfer Learning

Treinar uma CNN do zero exige milhões de imagens, o que costuma ser inviável em muitos projetos. O Transfer Learning aproveita redes pré-treinadas em grandes bases de dados, como a ImageNet, e adapta seus pesos para uma nova tarefa.

As primeiras camadas da rede já aprenderam padrões gerais (bordas, cores e texturas). No processo de ajuste fino (fine-tuning), essas características são refinadas para o domínio específico, neste caso, reconhecer escudos da Premier League.

D. Métricas de Avaliação

Para avaliar o desempenho do modelo, foram utilizadas métricas clássicas de classificação.

1) *Matriz de Confusão*: É uma tabela que compara as classes reais com as classes previstas.

Nela observamos:

- Verdadeiro Positivo (TP): acerto correto da classe.
- Falso Positivo (FP): o modelo previu errado uma classe.

- Falso Negativo (FN): a classe correta não foi reconhecida.

A matriz permite identificar quais classes são mais confundidas entre si.

2) *Acurácia*: Representa a proporção total de acertos em relação ao número de amostras.

3) Precisão e Revocação:

- Precisão: entre todas as previsões para uma classe, quantas estavam corretas.
- Revocação: entre todas as amostras reais daquela classe, quantas o modelo encontrou.

III. METODOLOGIA

A. Conjunto de Dados

O dataset foi construído contendo imagens dos 20 times da Premier League, totalizando aproximadamente 300 imagens por classe (150 sintéticas + 150 em contextos reais) que, após pré-processamento, foram divididas em conjuntos de Treinamento (80%) e Validação (20%). A divisão foi estratificada para garantir o balanceamento de classes. O conjunto é composto por:

- 1) Domínio Sintético: Escudos digitais planos em fundo branco ou transparente.
- 2) Domínio Real: Fotografias de torcedores, uniformes de jogo e mercadorias (*merchandising*), onde o escudo sofre rotações, oclusões e variações de iluminação.

B. Arquitetura e Transfer Learning

Adotou-se a arquitetura ResNet18 [1], escolhida pelo equilíbrio entre eficiência computacional e profundidade para extração de *features*. Utilizou-se a técnica de *Transfer Learning* com pesos pré-treinados na base ImageNet. Duas estratégias de treinamento foram comparadas:

- 1) *Feature Extraction (Baseline)*: Congelamento de todas as camadas convolucionais, treinando apenas o classificador linear final.
- 2) *Fine-Tuning (Proposta)*: Descongelamento de toda a rede, permitindo o ajuste fino dos pesos convolucionais para adaptar a extração de características à geometria específica dos escudos.

C. Estratégia de Data Augmentation

Para combater o viés de domínio observado no experimento baseline, aplicou-se um *pipeline* agressivo de aumento de dados utilizando a biblioteca Albumentations [3]. As transformações incluíram:

- Geométricas: Rotação (limit=45°), *Flip* horizontal e Perspectiva, simulando a visualização de escudos em tecidos dobrados e diferentes ângulos.
- Cromáticas e Textura: *ColorJitter* (alteração aleatória de brilho, contraste e saturação) e *GaussianBlur* (desfoque), forçando a rede a se concentrar na forma e nos elementos internos do escudo, desacoplando o aprendizado de características irrelevantes como luminosidade ou nitidez.
- Oclusão e Ruído: *CoarseDropout* (simulando obstruções parciais por objetos) e *GaussianNoise*, para reduzir a dependência de imagens limpas e de alta resolução.

D. Configuração de Treinamento

O modelo foi implementado utilizando o framework PyTorch. As imagens de entrada foram redimensionadas para 224 x 224 pixels, seguindo o padrão da arquitetura ResNet, e normalizadas com as médias e desvios-padrão da ImageNet. O treinamento foi realizado em lotes (*batch size*) de 32 imagens. Todo o processo experimental foi executado no ambiente Google Colab, utilizando uma GPU T4 para aceleração de hardware.

O modelo final foi treinado por 15 épocas, utilizando o otimizador Adam com uma taxa de aprendizado reduzida ($\eta = 10^{-4}$) para evitar a degradação dos pesos pré-treinados durante o *fine-tuning*. A função de perda utilizada foi a *CrossEntropyLoss*.

IV. RESULTADOS E DISCUSSÃO

A. Análise do Baseline vs. Modelo Proposto

O experimento inicial (Baseline), com pesos congelados e augmentation leve, atingiu uma acurácia de validação de 72%. A análise qualitativa dos erros revelou um forte viés de cor e fundo: o modelo classificava corretamente quase a totalidade das imagens sintéticas, mas falhava consistentemente em imagens reais onde a cor predominante do fundo divergia da cor primária do time (ex.: confundindo um gorro preto do Watford com o Crystal Palace).

Após a aplicação do *Fine-Tuning* e de um pipeline de *Data Augmentation* mais agressivo, o modelo proposto obteve uma acurácia de 88.2% na validação. O gráfico de convergência (Fig. 1) demonstra que o modelo estabilizou por volta da décima época, indicando um processo de aprendizagem consistente.

B. Capacidade de Generalização

Observou-se uma diferença (*gap*) inferior a 3% entre a acurácia de treino (91%) e a de validação (88%). Essa proximidade sugere que as técnicas de regularização utilizadas (Augmentation) foram eficazes em prevenir o *overfitting*, mesmo com um dataset limitado. O modelo demonstrou capacidade de reconhecer a geometria dos escudos independentemente do fundo, iluminação ou deformações, superando a limitação inicial associada a imagens sintéticas com fundo branco. A evolução da acurácia ao longo das épocas (Fig. 1) confirma esse comportamento, evidenciando convergência estável e ausência de sinais de overfitting significativo.

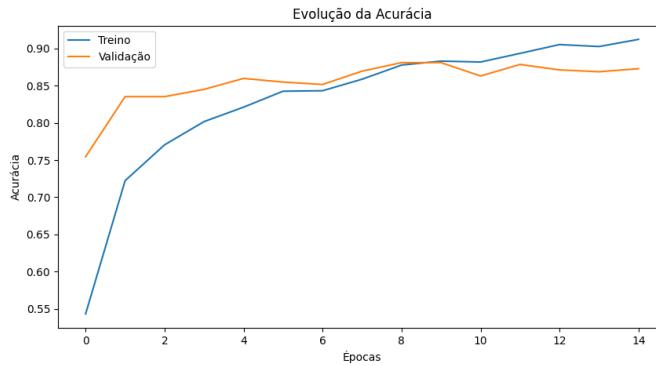


Fig. 1. Evolução da Acurácia

C. Matriz de Confusão e Análise dos Erros

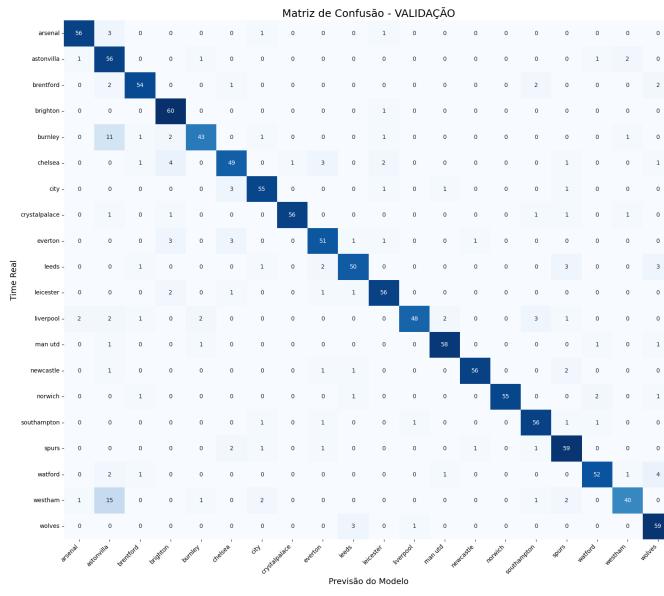


Fig. 2. Matriz de Confusão do conjunto de validação

A matriz de confusão (Fig. 2) evidencia que os erros remanescentes concentram-se, majoritariamente, em pares de times com escudos geometricamente similares ou paletas de cores muito próximas. Esse padrão contrasta com o Baseline, onde os erros tinham origem em vieses de fundo; no modelo final, os erros tornam-se visualmente plausíveis e estão relacionados à semelhança real entre os escudos.

O caso mais expressivo é a confusão entre West Ham e Aston Villa (Fig. 3), com 15 ocorrências, seguida por Burnley e Aston Villa, com 11 ocorrências. Os três clubes utilizam a tradicional paleta *claret & blue*, o que reduz significativamente o contraste entre as classes e dificulta a extração de texturas discriminativas pela rede. Além da paleta semelhante, West Ham e Villa possuem escudos com formato heráldico e elementos centrais dourados, contribuindo para uma sobreposição estrutural que favorece a confusão.

Além desses casos extremos, duas confusões adicionais apresentaram 4 ocorrências cada: Chelsea → Brighton e Watford → Wolves. Em ambos os pares, a sobreposição de cores dominantes e a presença de formas internas de baixo contraste

sugerem que, em condições de baixa resolução ou oclusão, a rede tende a capturar padrões cromáticos mais amplos em vez de detalhes específicos de cada emblema.

Outras confusões relevantes, embora menos frequentes (3 ocorrências cada), foram observadas entre pares como Chelsea → Everton, Everton → Chelsea, Arsenal → Aston Villa e Leeds → Spurs/Wolves. Em todos esses casos, padrões cromáticos semelhantes (como a predominância do azul nos escudos de Chelsea e Everton) ou estruturas visuais com alto nível de simetria levaram a erros recorrentes.

Confusões de baixo impacto (1 ou 2 ocorrências), como Chelsea–Leicester, Liverpool–Manchester United, Watford–Aston Villa e diversos outros pares, foram agrupadas e não discutidas individualmente por representarem casos estatisticamente pouco significativos. Esses erros tendem a refletir variabilidade natural das imagens, como sombras, oclusões parciais ou artefatos locais, e não padrões estruturais consistentes de falha do modelo.

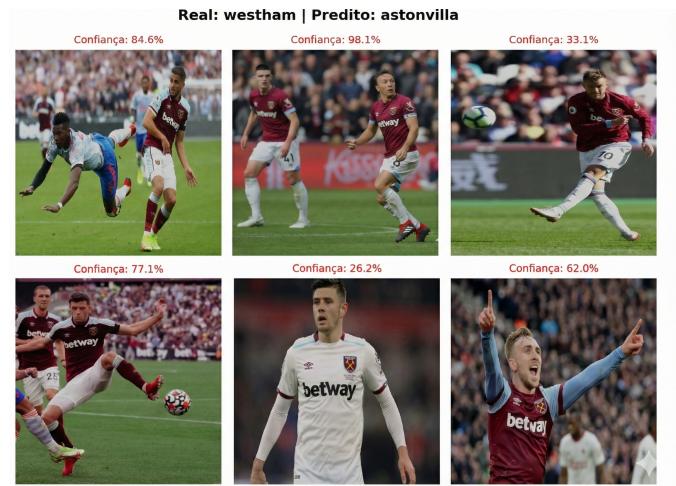


Fig. 3. Exemplos da confusão recorrente entre West Ham (rótulo real) e Aston Villa (predição do modelo). Em diversos cenários, incluindo baixa resolução, variação de pose e oclusão parcial, a paleta claret & blue e o formato semelhante dos escudos reduzem a discriminação visual, levando à classificação incorreta.

D. Métricas por Classe: Precisão, Recall e F1-Score

Além da matriz de confusão, foram calculadas as métricas clássicas de avaliação por classe: *Precisão* (*Precision*), *Recall* e *F1-Score*. Essas métricas permitem quantificar de forma mais detalhada o desempenho do modelo para cada equipe individualmente.

A precisão mede a proporção de previsões corretas entre todas as amostras atribuídas pelo modelo a uma determinada classe:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (1)$$

O recall quantifica a capacidade do modelo de identificar corretamente as instâncias reais da classe:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2)$$

O F1-Score combina ambas as métricas por meio da média harmônica, equilibrando precisão e recall:

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

A Tabela I apresenta os valores obtidos para cada time no conjunto de validação. Observa-se que grande parte das classes apresenta desempenho elevado (F1 acima de 0.85), especialmente equipes com escudos altamente distintivos, como *Brighton*, *Norwich* e *Crystal Palace*, que alcançaram índices superiores a 0.95.

Por outro lado, classes que compartilham características visuais semelhantes, particularmente a paleta *claret & blue*, apresentaram valores reduzidos. O caso mais crítico é *West Ham*, com *recall* de apenas 0.58, refletindo sua alta taxa de confusão com *Aston Villa*, já destacada na matriz de confusão. Esse comportamento reforça que a similaridade cromática e estrutural entre determinados escudos interfere diretamente na capacidade discriminativa do modelo.

Tabela I
PRECISÃO, RECALL E F1-SCORE POR CLASSE

| Time | Precision | Recall | F1-Score |
|-------------------------|-----------|--------|----------|
| Arsenal | 0.81 | 0.93 | 0.87 |
| Aston Villa | 0.88 | 0.87 | 0.88 |
| Brentford | 0.87 | 0.89 | 0.88 |
| Brighton | 0.97 | 0.97 | 0.97 |
| Burnley | 0.76 | 0.90 | 0.82 |
| Chelsea | 0.78 | 0.87 | 0.82 |
| City | 0.90 | 0.93 | 0.92 |
| Crystal Palace | 0.95 | 0.95 | 0.95 |
| Everton | 0.86 | 0.73 | 0.79 |
| Leeds | 0.84 | 0.78 | 0.81 |
| Leicester | 0.85 | 0.92 | 0.88 |
| Liverpool | 0.89 | 0.89 | 0.89 |
| Man Utd | 1.00 | 0.87 | 0.93 |
| Newcastle | 0.97 | 0.92 | 0.94 |
| Norwich | 0.98 | 0.95 | 0.97 |
| Southampton | 0.85 | 0.90 | 0.87 |
| Spurs | 0.78 | 0.91 | 0.84 |
| Watford | 0.94 | 0.80 | 0.87 |
| West Ham | 0.88 | 0.58 | 0.70 |
| Wolves | 0.86 | 0.95 | 0.90 |
| Média (Accuracy) | 0.88 | 0.88 | 0.88 |
| Macro Avg | 0.88 | 0.88 | 0.87 |
| Weighted Avg | 0.88 | 0.88 | 0.87 |

V. CONCLUSÃO

Este trabalho demonstrou que a aplicação direta de Transfer Learning com pesos congelados é insuficiente para tarefas de classificação sujeitas a alta variabilidade de domínio, como no caso de escudos da Premier League em ambientes reais. O modelo baseline apresentou forte dependência do fundo e da paleta de cores, atingindo apenas 72% de acurácia na validação, apesar de performar quase perfeitamente em imagens sintéticas.

A combinação de Fine-Tuning completo com um pipeline agressivo de Data Augmentation mostrou-se essencial para superar esse viés e permitir que a rede aprendesse representações mais invariantes. A acurácia final atingiu 88%, com gap treino–validação inferior a 3%, indicando boa capacidade de

generalização mesmo diante de um conjunto reduzido de imagens reais.

A análise qualitativa evidenciou que os erros remanescentes concentraram-se em classes intrinsecamente semelhantes (como *Aston Villa*, *West Ham* e *Burnley*), enquanto classes mais distintas apresentaram métricas próximas de 95-98% de F1-Score. Além disso, o uso de técnicas como CoarseDropout, ruído gaussiano e transformações geométricas contribuiu diretamente para que o modelo deixasse de depender de pistas superficiais (como o fundo branco) e passasse a reconhecer padrões estruturais dos escudos.

Conclui-se que perturbar intencionalmente as imagens durante o treinamento é uma estratégia eficaz para aumentar a robustez de modelos de visão computacional em cenários de escassez de dados e forte discrepância visual entre treino e inferência. Como trabalhos futuros, sugere-se explorar arquiteturas mais profundas (ResNet50, EfficientNet), técnicas de mixup / cutmix, e estratégias de adaptação de domínio para reduzir ainda mais os erros em classes com escudos similares.

REFERENCES

- [1] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, 2016, pp. 770–778.
- [2] O. Russakovsky *et al.*, “ImageNet large scale visual recognition challenge,” *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, 2015.
- [3] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, and A. A. Kalinin, “Albumentations: Fast and flexible image augmentations,” *Information*, vol. 11, no. 2, p. 125, 2020.
- [4] S. J. Pan and Q. Yang, “A survey on transfer learning,” *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
- [5] A. Paszke *et al.*, “PyTorch: An imperative style, high-performance deep learning library,” in *Adv. Neural Inf. Process. Syst.*, 2019, pp. 8024–8035.
- [6] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” in *Proc. Int. Conf. Learn. Representations (ICLR)*, 2015.