DOI: 10.5748/19CONTECSI/REX//6959

# ANÁLISE DE ALGORITMOS DE RECONHECIMENTO FACIAL: HOG (HISTOGRAM OF ORIENTED GRADIENTS) E YOLO (YOU ONLY LOOK ONCE)

**Dilermando Piva Jr** ; https://orcid.org/0000-0002-2534-9618
Fatec Itu / Fatec Sorocaba

**Erick José da Silva** ; https://orcid.org/0000-0001-9819-4912
Fatec Itu

# ANÁLISE DE ALGORITMOS DE RECONHECIMENTO FACIAL: HOG (HISTOGRAM OF ORIENTED GRADIENTS) E YOLO (YOU ONLY LOOK ONCE)

# ANALYSIS OF FACIAL RECOGNITION ALGORITHMS: HOG (HISTOGRAM OF ORIENTED GRADIENTS) AND YOLO (YOU ONLY LOOK ONCE)

**RESUMO:** Com o crescente uso de aplicações voltadas ao reconhecimento facial, tornou-se fundamental analisar quais tecnologias são capazes de detectar facilmente pessoas sem depender de um elevado tempo de processamento. Além disso, modelos desenvolvidos com Inteligência Artificial (IA) vêm se tornando cada vez mais sofisticados, permitindo que sejam criados sistemas de reconhecimento facial baseados em aprendizagem profunda. Isso foi possível graças ao uso de computadores com alto poder de processamento, que utilizam GPUs em sua arquitetura. Tendo em vista isso, o presente trabalho é um estudo comparativo dos algoritmos de reconhecimento facial HOG (Histogram of Oriented Gradients) e YOLO (You Only Look Once), com foco em avaliar os seus desempenhos através da acurácia e da latência. Para atingir o objetivo proposto, utilizou-se os dois algoritmos para analisar 1000 fotos, oriundas das bases de imagens Yalefaces, Labeled Faces in the Wild (LFW) e Dogs VS Cats. Desse conjunto, 600 imagens eram de pessoas. Os resultados obtidos foram que HOG possui uma maior acurácia que YOLO. Todavia, YOLO apresenta um menor tempo de processamento, atingindo o valor máximo de 0,65 segundos para processar uma imagem. Através desse estudo, espera-se mostrar as vantagens e desvantagens de cada um dos algoritmos, bem como as características que os distinguem entre si.

**ABSTRACT:** With the growing use of applications aimed at facial recognition, it has become essential to analyze which technologies are able to easily detect people without depending on a high processing time. In addition, models developed with Artificial Intelligence (AI) are becoming increasingly sophisticated, allowing deep learning-based facial recognition systems to be created. This was possible thanks to the use of computers with high processing power, which use GPUs in their architecture. In view of this, the present work is a comparative study of the facial recognition algorithms HOG (Histogram of Oriented Gradients) and YOLO (You Only Look Once), focusing on evaluating their performance through accuracy and latency. To achieve the proposed objective, both algorithms were used to analyze 1000 photos from the Yalefaces, Labeled Faces in the Wild (LFW) and Dogs VS Cats image databases. Of this set, 600 images were of people. The results obtained were that HOG has a higher accuracy than YOLO. However, YOLO has a shorter processing time, reaching a maximum value of 0.65 seconds to process an image. Through this study, it is expected to show the advantages and disadvantages of each of the algorithms, as well as the characteristics that distinguish them from each other.

**PALAVRAS-CHAVE:** Reconhecimento facial. HOG. YOLO. Aprendizagem Profunda. Machine Learning.

**KEYWORD:** Facial recognition. HOG. YOLO. Deep Learning. Machine Learning.

## 1 INTRODUÇÃO

Desde os primeiros trabalhos de reconhecimento facial por computadores, a tecnologia evoluiu a passos largos. Muitas técnicas de processamento de imagem foram criadas com o objetivo de melhorar a detecção de rostos e objetos, gerando avanços importantes na área de Inteligência Artificial (IA). Na década de 2000, descritores de recursos foram criados com o foco em reduzir o tempo de processamento sem perder a eficácia de extrair as características da imagem. Além disso, com a evolução computacional, foi possível aplicar em dispositivos móveis algoritmos classificadores de objetos, tornando acessível ao público utilizar soluções baseadas em *Machine Learning*, como a possibilidade de câmeras identificarem rostos e até mesmo *smartphones* pesquisarem em mecanismos de buscas com base em fotos tiradas pelo usuário.

Considerando os importantes avanços da IA em reconhecimento facial, o presente trabalho tem como objetivo mostrar dois dos mais conhecidos algoritmos utilizados nessa área: *Histogram of Oriented Gradients* (HOG) e *You Only Look Once* (YOLO). Serão explicados o funcionamento e as características desses dois algoritmos, sendo que no final do relatório haverá uma avaliação da *perfomance* deles. Os resultados dessa análise baseiam-se na quantidade de acertos que HOG e YOLO tiveram ao verificar a presença de rostos em uma amostra de 1000 imagens, oriundas das bases Yalefaces, Labeled Faces in the Wild (LFW) e Dogs VS Cats.

## 2 METODOLOGIA

HOG e YOLO tiveram o seu desempenho avaliado por meio da acurácia e do tempo de processamento. Para obter essas duas métricas, foram utilizadas três bases de imagens: Yalefaces, *Labeled Faces in the Wild* (LFW) e *Dogs VS Cats*. As duas primeiras bases possuem imagens de pessoas, enquanto a última contém fotos de cães e gatos. Desses três bancos, se retirou uma amostra de 1000 fotos, que foram colocadas em um único diretório. Dessas 1000 imagens, 600 foram de pessoas.

As especificações do hardware utilizado para os testes foram as seguintes: processador AMD-Ryzen 5 1600 AF; placa-mãe Asus B450M Gaming; memória RAM 16 GB 3200 MHz DDR4; SSD NVMe 256 GB; placa de vídeo RX 580 4GB.

## 3 DESENVOLVIMENTO

### 3.1 HOG (*Histogram of Oriented Gradients*)

Desenvolvido por Dalal e Triggs, o Histograma de Gradientes Orientados (ou HOG) é um descritor de recursos cujo diferencial está em sua arquitetura, que foi criada para a identificação de pessoas (DALAL; TRIGGS, 2005). O seu funcionamento se dá através da distribuição (histograma) de gradientes orientados de uma imagem. Por meio desses atributos, HOG é capaz de extrair bordas de um ou mais objetos, para depois enviá-las a um algoritmo classificador responsável por definir se na imagem analisada há ou não uma face.

Gradientes representam como a tonalidade da cor varia no decorrer da foto, seguindo uma determinada orientação. Quando os *pixels* de uma região apresentam uma mudança abrupta de cor, significa que aquele trecho analisado é a borda de um objeto. Em termos matemáticos, o gradiente é obtido através da derivada de uma função multivariável.

Conforme Dalal e Triggs comentam no artigo, o uso de gradientes orientados na detecção de objetos não foi algo novo que apareceu em HOG. Muitos trabalhos anteriores já utilizavam essas características das imagens para detectar objetos e humanos. McConnell (1985) patenteou um descritor baseado em gradientes. Freeman e Roth (1995) desenvolveram um algoritmo que captava movimentos das mãos através desses atributos. Um ano depois, Freeman *et al* (1996) aplicaram esse trabalho em jogos digitais. O objetivo desses pesquisadores era permitir a interação do usuário com os jogos através de movimentos das mãos e do corpo, sem precisar de *joysticks*.

Mais um aspecto que HOG compartilha com outros descritores, em particular com o proposto por Belongie *et al* (2001), é o contexto de forma (*Shape Context*). O algoritmo de Belongie divide a imagem em blocos para conseguir extrair as bordas de um objeto. *Shape Context* não adota gradientes orientados para realizar esta tarefa. Dalal e Triggs se basearam no trabalho de Belongie em dividir os *pixels* de uma imagem em quadros e aplicaram um histograma de gradientes orientados em cada célula, distinguindo HOG dos demais descritores.

Ao comparar o desempenho de HOG com *Shape Context* e os demais algoritmos de gradientes orientados, Dalal e Triggs afirmam que o seu descritor retorna menos falsos positivos na identificação de humanos. Isso porque os detectores de ponto chave anteriores ao HOG não apresentam uma estrutura voltada a identificação de pessoas.

Dalal e Triggs utilizaram o *Support Vector Machines* (SVM) como algoritmo classificador. Raschka (2016) explica que o SVM tem como objetivo encontrar, dentro de um conjunto de dados, a maior margem de classificação. Define-se como margem a distância entre os agrupamentos de

dados, que são construídos com base em padrões obtidos nos dados (RASCHKA, 2016). Ao encontrar essa região, o SVM é capaz de classificar se um valor pertence a um grupo X ou a um Y.

### 3.2 YOLO (*You Only Look Once*)

Lançado em 2016 por Redmon *et al* (2016), YOLO (*You Only Look Once*) é uma rede neural profunda, classificada como detectora de objetos. Sua principal função não é apenas reconhecer qual objeto está presente na imagem (como ocorre em algoritmos classificadores), mas também identificar em qual posição ele está localizado. Em outras palavras, YOLO é capaz de responder duas perguntas: quais objetos estão na imagem? Onde eles estão posicionados?

A detecção é formada por dois processos: a classificação e a localização. Algoritmos classificadores são capazes de atribuir um rótulo a um conjunto de recursos extraídos da imagem (FORSYTH; PONCE, 2012). Já os localizadores distinguem o objeto do pano de fundo, e identificam as relações espaciais entre os objetos na imagem (BLASCHKO; LAMPERT, 2018). YOLO é a combinação das características presentes nesses algoritmos, uma vez que utiliza redes neurais profundas para reconhecer os objetos. Além disso, esses dois processos juntos permitem identificar vários objetos diferentes em uma imagem.

Outra característica importante de YOLO é ser um detector de objetos em tempo real. Redmon *et al* (2016) fizeram testes que comprovaram um ótimo desempenho do algoritmo ao reconhecer objetos através de *webcams*. O tempo de processamento foi rápido, chegando a atingir um valor menor que 25 milissegundos de latência. Com base nos testes feitos com *webcams*, os autores perceberam que YOLO apresentou uma solução eficaz e de fácil usabilidade. A detecção ocorria independente da movimentação do objeto ou da forma como a sua aparência mudava no vídeo.

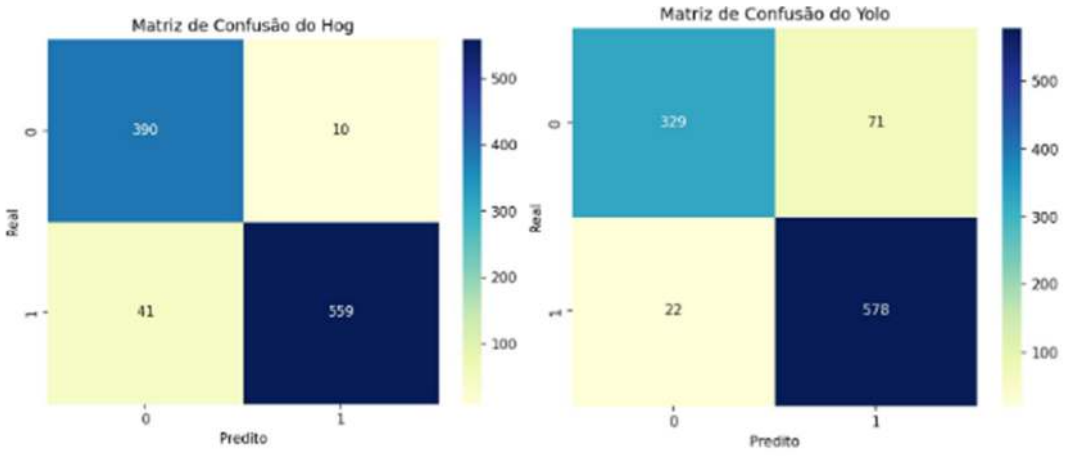
O modelo desenvolvido por Redmon *et al* (2016) é uma rede neural convolucional (CNN) *open source*, chamada Darknet. Suas primeiras camadas são responsáveis por extrair os recursos da imagem, enquanto as camadas totalmente conectadas preveem a probabilidade das classes e as coordenadas de saída. A rede possui 24 camadas convolucionais e duas totalmente conectadas. Além disso, cada camada possui uma redução 1 X 1 de seu antecessor, seguidas por camadas convolucionais 3 X 3.

Quanto as etapas que formam YOLO, inicialmente se aplica uma grade de células S x S sobre a imagem. Em seguida, a classificação e a localização dos objetos ocorrem de maneira simultânea. Por fim, a saída obtida é a detecção, com as caixas delimitadoras e os rótulos.

## 4 RESULTADOS OBTIDOS

Após realizar os testes com os dois algoritmos, uma série de informações foram geradas e com elas foram elaborados três gráficos comparativos entre os algoritmos HOG e YOLO. Na Figura 1 são evidenciadas as matrizes de confusão obtidas como resultados da avaliação. Esses visuais têm como objetivo comparar o número de resultados preditos pelo algoritmo com a quantidade real de dados por categoria. A matriz $M_{linha,coluna}$ divide os dados em quatro categorias: falsos positivos ($M_{0,1}$), verdadeiros positivos ($M_{1,1}$), falsos negativos ($M_{1,0}$) e verdadeiros negativos ($M_{0,0}$) (SCIKITLEARN, 2022).

**Figura 1 – Matrizes de confusão de HOG e YOLO.**




**Fonte: Elaborada pelos próprios autores.**

Além disso, com a matriz de confusão, é possível calcular a acurácia do algoritmo. Mishra (2018) explica que a acurácia de um modelo é dada pela razão entre a quantidade de acertos do algoritmo e o total de predições realizadas. Ou seja:

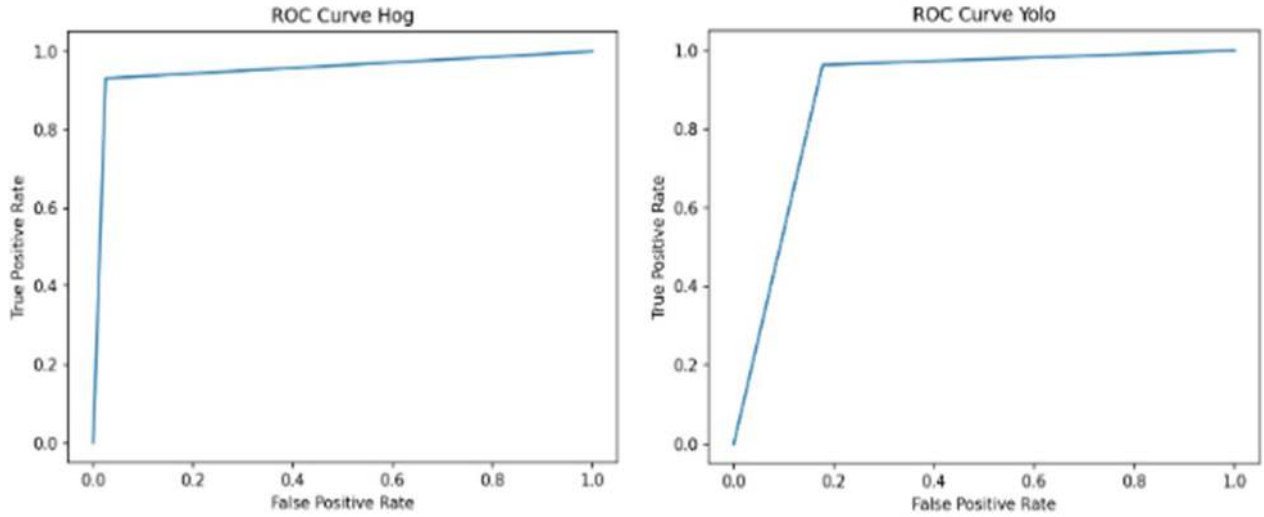
$$Acurácia = \frac{VP + VN}{VP + VN + FP + FN}$$

Onde: **VP** – verdeiros positivos; **VN** – verdadeiros negativos; **FP** – falsos positivos e **FN** – falsos negativos. Com base na matriz de confusão, HOG e YOLO possuem acurácias de 95% e 91%, aproximadamente, pois:

| | |
|---|---|
| $Acurácia_{HOG} = \frac{559 + 390}{559 + 390 + 41 + 10} = 0{,}949$ | $Acurácia_{YOLO} = \frac{578 + 329}{578 + 329 + 22 + 71} = 0{,}907$ |

A Figura 2 mostra a curva Característica de Operação do Receptor (ou ROC – *Receiver Operating Characteristic curve*) dos algoritmos. Segundo Avelar (2019): "ROC é uma curva de probabilidade. Ela é criada traçando a taxa de verdadeiros positivos contra a taxa de falsos-positivos. Ou seja, número de vezes que o classificador acertou a predição contra o número de vezes que o classificador errou a predição". Geralmente, o eixo X da curva é a taxa de falsos positivos, que é dada por falsos positivos/(falsos positivos + verdadeiros negativos), enquanto o eixo Y é gerado pela taxa de verdadeiros positivos, cujo cálculo é verdadeiros positivos/(verdadeiros positivos + falsos negativos) (AVELAR 2019).

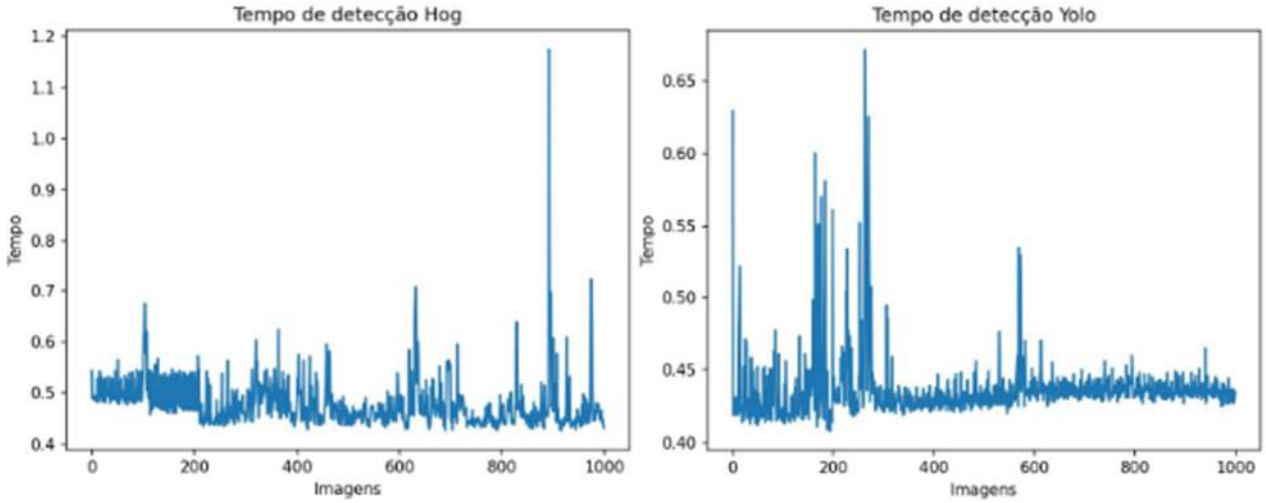
**Figura 2 – Curva ROC de HOG e YOLO.**




**Fonte: Elaborada pelos próprios autores.**

Por fim, a Figura 3 mostra o tempo de detecção dos algoritmos. O gráfico mostra o tempo em função da quantidade de imagens.

**Figura 3 – Tempo de detecção de HOG e YOLO.**




**Fonte: Elaborada pelos próprios autores.**

## 5 CONSIDERAÇÕES FINAIS

O presente trabalho evidenciou as principais características do HOG e do YOLO, algoritmos utilizados nas áreas de detecção e reconhecimento facial. Também foi realizada uma análise comparativa, mostrando os seus desempenhos. Por meio dos resultados obtidos nesse estudo, conclui-se que HOG apresentou maior acurácia que YOLO, porém com tempo de detecção maior. HOG teve uma acurácia 4,2% maior, enquanto YOLO foi em média 22% mais rápido. Isso corrobora com Redmon *et al* (2016), que definem YOLO como um detector de objetos em tempo real devido a sua menor latência.

Para trabalhos futuros, seria interessante avaliar a *perfomance* de todas as versões de YOLO, bem como realizar uma análise comparativa com outras redes neurais utilizadas em reconhecimento facial. Quanto ao HOG, compará-lo com outros descritores de recursos e algoritmos de *Machine Learning* que não utilizam o paradigma de redes neurais. Por fim, um trabalho mostrando o desempenho de HOG e YOLO na detecção de emoções também seria relevante para a área de reconhecimento facial.

## REFERÊNCIAS

AVELAR, Adriano. **O que é AUC e ROC nos modelos de *Machine Learning***. 2019. Disponível em: https://medium.com/@eam.avelar/o-que-%C3%A9-auc-e-roc-nos-modelos-de-machine-learning-2e2c4112033d . Acesso em: 26 jun. 2022.

BELONGIE, Serge; MALIK, Jitendra; PUZICHA, Jan. *Matching Shapes*. **8th IEEE International Conference on Computer Vision**, Vancouver, Canadá, p. 454-461, jul. 2001.

BLASCHKO, Matthew B.; LAMPERT, Christoph H.. *Learning to Localize Objects with Structured Output Regression*. **Lecture Notes in Computer Science**, [S.L.], p. 2-15, 2008. Springer Berlin Heidelberg. http://dx.doi.org/10.1007/978-3-540-88682-2_2 .

DALAL, N.; TRIGGS, B.. *Histograms of Oriented Gradients for Human Detection*. **2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)**, [S.L.], p. 886-893, 2005. IEEE. http://dx.doi.org/10.1109/cvpr.2005.177 .

FORSYTH, David A.; PONCE, Jean. ***Computer Vision:*** *a modern approach*. 2. ed. New Jersey: Pearson Education, 2012. 761 p.

FREEMAN, W.T. *et al*. *Computer vision for computer games*. **Proceedings of the Second International Conference on Automatic Face and Gesture Recognition**, [S.L.], p. 100-105, 1996. IEEE Comput. Soc. Press. http://dx.doi.org/10.1109/afgr.1996.557250 .

FREEMAN, William T.; ROTH, Michal. *Orientation Histograms for Hand Gesture Recognition*. **IEEE Intl. Wkshp. on Automatic Face and Gesture Recognition**, Zurique, Suíça, p. 296-301, jun. 1995.

MCCONNELL, R. K.. ***Method of and apparatus for pattern recognition***. . US n. 4567610. Depósito: 22 jul. 1982. Concessão: 28 jan. 1986.

MISHRA, Aditya. ***Metrics to Evaluate your Machine Learning Algorithm***. 2018. Disponível em: https://towardsdatascience.com/metrics-to-evaluate-your-machine-learning-algorithm-f10ba6e38234 . Acesso em: 26 jun. 2022.

RASCHKA, Sebastian. ***Python Machine Learning***. Birmingham, Inglaterra: Packt Publishing Ltd, 2016. 425 p.

REDMON, Joseph *et al*. *You Only Look Once: unified, real-time object detection*. **2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**, [S.L.], p. 779-788, jun. 2016. IEEE. http://dx.doi.org/10.1109/cvpr.2016.91 .

SCIKITLEARN. **sklearn.metrics.confusion_matrix**. Disponível em: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.confusion_matrix.html Acesso em: 26 jun. 2022.