

# Visão Computacional: reconhecimento de texto com OCR e OpenCV

[cursos.alura.com.br/course/visao-computacional-reconhecimento-texto-ocr-opencv/task/113548](https://cursos.alura.com.br/course/visao-computacional-reconhecimento-texto-ocr-opencv/task/113548)

Os modos de segmentação de página, mais conhecidos como PSM, são extremamente importantes para melhorar a precisão de acertos usando OCR, principalmente quando temos imagens mais complexas. Nesses casos, o Tesseract OCR pode ter dois retornos diferentes do previsto:

- Retorna um resultado vazio: não é possível reconhecer nenhum texto na imagem com o OCR;
- Retorna um resultado com caracteres sem sentido: tenta fazer o reconhecimento de texto com o OCR, mas está totalmente incorreto.

Para listar os 14 PSMs do Tesseract basta passarmos o código `!tesseract --help-psm`

E o resultado obtido será:

Page segmentation modes:

- 0 Orientation and script detection (OSD) only.
- 1 Automatic page segmentation with OSD.
- 2 Automatic page segmentation, but no OSD, or OCR.
- 3 Fully automatic page segmentation, but no OSD. (Default)
- 4 Assume a single column of text of variable sizes.
- 5 Assume a single uniform block of vertically aligned text.
- 6 Assume a single uniform block of text.
- 7 Treat the image as a single text line.
- 8 Treat the image as a single word.
- 9 Treat the image as a single word in a circle.
- 10 Treat the image as a single character.
- 11 Sparse text. Find as much text as possible in no particular order.
- 12 Sparse text with OSD.
- 13 Raw line. Treat the image as a single text line, bypassing hacks that are Tesseract-specific.

Vamos entender o que cada um dos PSMs significa.

- **PSM 0 - Apenas orientação e detecção de script (OSD):** Esse modo não executa OCR, ele apenas nos dá informações sobre a imagem: como a imagem está disposta em graus (se está rotacionada ou não), a confiança da escrita (sinais gráficos/sistemas de escrita) como por exemplo se é latim, han, cirílico, etc.
- **PSM 1 - Segmentação automática de página com OSD:** Esse modo não executa OCR e não retorna um resultado do OSD, mas ele ajusta o OSD de forma automática.
- **PSM 2 - Segmentação de página automática, mas sem OSD ou OCR:** Faz apenas a segmentação da página, sem retirar informações de OSD ou textos com OCR.

- **PSM 3 - Segmentação de página totalmente automática, mas sem OSD:** Modo padrão do Tesseract OCR, tenta encaixar o PSM correto no texto que está na imagem, mas sem retorno de OSD.
- **PSM 4 - Assuma uma única coluna de texto de tamanhos variáveis:** O OCR nesse caso lê a imagem como uma coluna, linha a linha, mesmo com textos de diferentes tamanhos. Isso pode ser aplicado por exemplo em dados de planilhas, tabelas ou recibos.
- **PSM 5 - Assuma um único bloco uniforme de texto alinhado verticalmente:** Nesse caso de PSM ao invés de termos a leitura da imagem na horizontal, temos ela na vertical, com as linhas em modo vertical, como se a imagem, a mesma que podemos usar no PSM 4, estivesse rotacionada 90°. PSM 5 pode ser aplicado em dados de planilhas, tabelas, recibos, entre outros que tenham blocos de textos em linhas verticais.
- **PSM 6 - Assuma um único bloco uniforme de texto:** Esse modo de segmentação pode ser utilizado para textos como páginas de livros, por exemplo, que tem uma única fonte. Nesses casos, quando o texto é uma única fonte sem qualquer variação, temos um texto uniforme e simples para o Tesseract compreender.
- **PSM 7 - Trate a imagem como uma única linha de texto:** Esse modo é utilizado quando trabalhamos com uma única linha de texto uniforme, como por exemplo placas de carro.
- **PSM 8 - Trate a imagem como uma única palavra:** Esse modo é utilizado apenas quando temos uma única palavra, um exemplo de caso é em placas como placas de trânsito “PARE” ou placas de alerta com “SAÍDA”.
- **PSM 9 - Trate a imagem como uma única palavra em um círculo:** O PSM 9 pode ser utilizado em dois casos específicos:

1 - quando o texto está realmente dentro de um círculo;

2 - quando o texto está em torno de um círculo invisível, ou seja, o texto está em formato de arco.

- **PSM 10 - Trate a imagem como um único caractere:** Esse modo deve ser utilizado quando o tratamento da imagem já foi feito, a extração de cada caractere individual da imagem já foi realizada e o PSM é utilizado para reconhecimento do caractere.
- **PSM 11 - Encontre o máximo de texto possível em nenhuma ordem específica:** A detecção de texto esparsa pode ser útil quando há muito texto em uma imagem que você precisa extrair. Ao usar este modo, você normalmente não se importa com a ordem ou agrupamento do texto, mas sim com o próprio texto.
- **PSM 12 - Texto esparsa com orientação e detecção de script:** O PSM 12 é semelhante ao PSM 0 (não executa OCR, apenas nos dá informações sobre a imagem), mas é utilizado em texto esparsa como o PSM 11.

- **PSM 13 - Trate a imagem como uma única linha de texto, ignorando hacks específicos do Tesseract:** Quando o resultado final de uma imagem com uma única linha de texto é nenhum texto detectado (seja por termos uma fonte que o Tesseract não reconhece, ou o texto é recortado ou até mesmo estilizado de alguma forma) é possível testar o PSM 13 que trata o texto como uma única linha bruta de texto.

Todos os modos de PSM estão disponíveis para uso no Tesseract OCR, e caso o retorno seja algum dos dois que vimos no início, mesmo com o uso de um PSM específico, o recomendado é a troca de PSM.

Discutir no Fórum Próxima Atividade