

# Relatório de Análise exploratória

Hugo Henrique Correia da Silva - 25 de maio de 2025

## VISÃO GERAL

O presente relatório tem como objetivo apresentar a seleção, organização e análise inicial do MPII Human Pose Dataset, que foi selecionado como foco de estudo devido à sua relevância na área de visão computacional, especialmente nas tarefas de estimativa de pose humana e reconhecimento de ações. Esta base de dados é amplamente utilizada na literatura científica e representa um benchmark importante para o desenvolvimento e validação de modelos de aprendizado profundo.

A análise exploratória aqui conduzida busca levantar considerações iniciais sobre as características dos dados selecionados, tanto no que se refere às imagens disponíveis, quanto às anotações fornecidas pelo arquivo de metadados. As informações contidas neste arquivo incluem, entre outros, os pontos-chave (*keypoints*) das articulações humanas, a visibilidade de cada ponto, a ação associada à imagem, e a indicação de qual indivíduo é o principal em cenas com múltiplas pessoas.

## OBJETIVOS

1. Seleção e organização de dados
2. Análise descritiva do conjunto criado

## Dataset

A base é composta por aproximadamente 25 mil imagens estáticas extraídas de vídeos do YouTube, totalizando mais de 40 mil pessoas anotadas. Para cada pessoa identificada, são fornecidas anotações detalhadas dos 16 pontos-chave do corpo humano, incluindo membros superiores e inferiores, tronco e cabeça. Além das coordenadas bidimensionais de cada ponto, também é fornecida a informação sobre sua visibilidade, indicando se o ponto está visível ou oculto na imagem.

Para esse estudo, foram selecionadas 3 classes ao acaso, no qual foram selecionadas 300 imagens de cada classe, o caso registrado foi para as classes *dancing*, *miscellaneous* e *sports*.

## Análise da base

Inicialmente a exportação inicial da base de anotações que contém os metadados das imagens, nos trouxe uma distribuição de 18033 imagens que são distribuídas na tabela a seguir em suas respectivas classes.

Rótulo	Quantidade
sports	3593
inactivity quiet/light	189
miscellaneous	755
occupation	2412
water activities	1066
home activities	1416
lawn and garden	1130
religious activities	124
winter activities	778
conditioning exercise	1755
bicycling	533
fishing and hunting	725
dancing	722
walking	608
running	316
self care	90
home repair	1090
volunteer activities	55
music playing	583
transportation	93

Cada registro desse tem informações relevantes acerca da imagem, como o número de pessoas na foto, nome da imagem que essa informação se refere, dimensão da foto, classe a que ela pertence, os *keypoints* (pontos mapeados do corpo humano) e se esses *keypoints* estão visíveis ou não e muitas outras informações. A seguir, temos um registro dentre muito outros

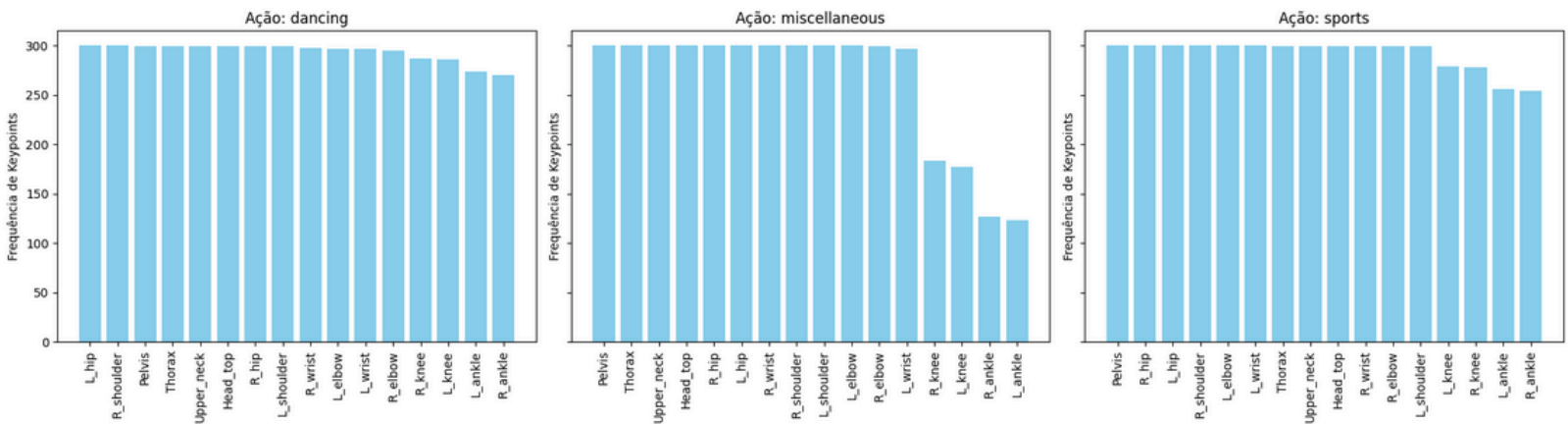
```
Pessoa 0: 16 keypoints
[{'id': 6, 'x': 438.0, 'y': 342.0, 'visible': 0}, {'id': 7, 'x': 398.0, 'y': 321.0, 'visible': 0}, {'id': 8, 'x': 400.3144, 'y': 322.4026, 'visible': -1}, {'id': 9, 'x': 362.6856, 'y': 299.5974, 'visible': -1}, {'id': 0, 'x': 414.0, 'y': 441.0, 'visible': 0}, {'id': 1, 'x': 400.0, 'y': 395.0, 'visible': 1}, {'id': 2, 'x': 423.0, 'y': 344.0, 'visible': 0}, {'id': 3, 'x': 453.0, 'y': 339.0, 'visible': 1}, {'id': 4, 'x': 459.0, 'y': 404.0, 'visible': 1}, {'id': 5, 'x': 488.0, 'y': 401.0, 'visible': 1}, {'id': 10, 'x': 343.0, 'y': 362.0, 'visible': 1}, {'id': 11, 'x': 367.0, 'y': 340.0, 'visible': 1}, {'id': 12, 'x': 384.0, 'y': 314.0, 'visible': 0}, {'id': 13, 'x': 411.0, 'y': 327.0, 'visible': 1}, {'id': 14, 'x': 423.0, 'y': 366.0, 'visible': 1}, {'id': 15, 'x': 403.0, 'y': 411.0, 'visible': 1}]
```

Os *keypoints*, são numerados de 0 a 15 e é esperado que todas as fotos tenham, para que possam ser utilizados para treinamento posteriormente. Entretanto, na divisão e seleção da base experimental, vemos na tabela a seguir que nem todos os dados tem todas as informações referente a *keypoints*, a classe *miscellaneous* como esperado é que menos tem informação com média de 14 pontos registrados, por ser uma classe de posições diversas algumas imagens devem ser difíceis de discernir a sua posição.

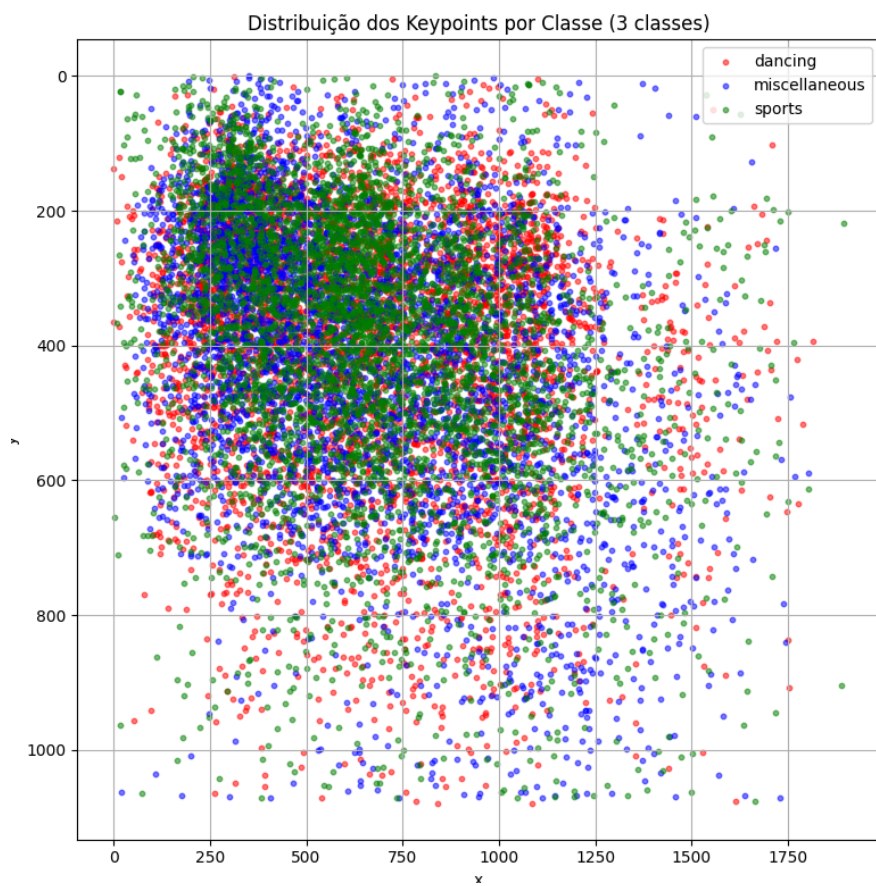
]:

	count	mean	std	min	25%	50%	75%	max
action								
dancing	300.0	15.660000	1.030180	10.0	16.0	16.0	16.0	16.0
miscellaneous	300.0	14.020000	1.808453	11.0	12.0	14.0	16.0	16.0
sports	300.0	15.536667	1.142954	11.0	16.0	16.0	16.0	16.0

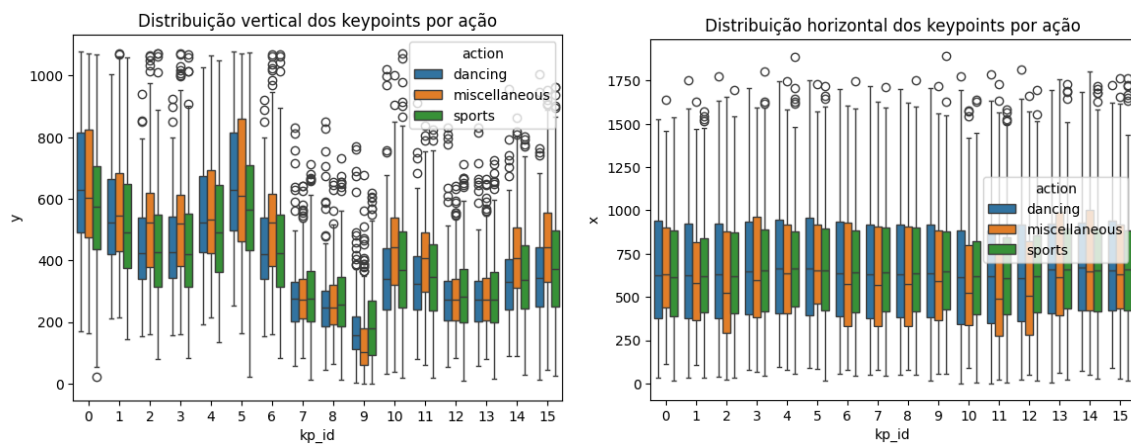
Ao vermos nas tabelas a seguir, conseguimos notar que os joelhos e os calcanhares, são os mais difíceis de identificar durante o processo de rotulagem, já que temos uma má distribuição considerável em relação aos outros *keypoints*.



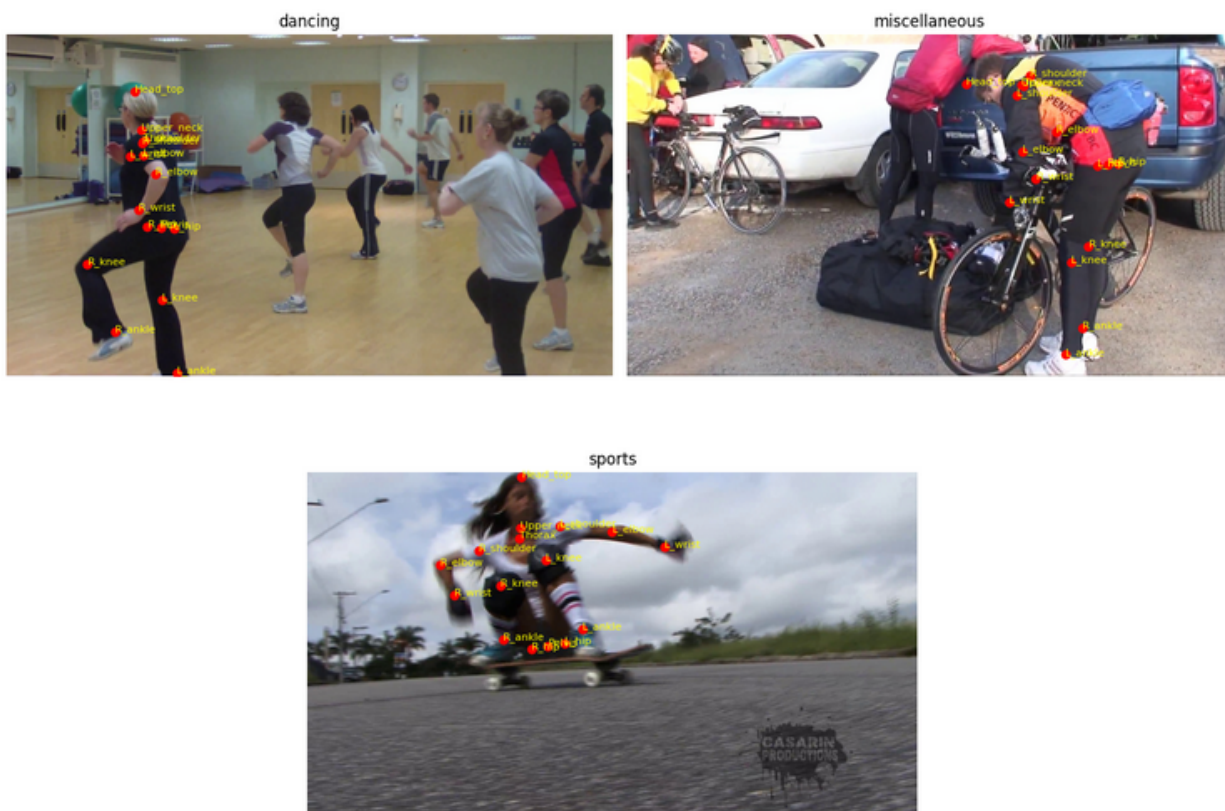
O próximo gráfico mostra os locais em que se encontram os *keypoints* em relação ao tamanho das imagens. O dataset tem imagens de tamanhos variáveis, mas colocando em plano conseguimos ver uma concentração dos pontos, na posição esquerda e superior do gráfico, outra coisa que notamos é que as classes se encontram bastante misturadas, revelando que não tem uma relação espacial a ser explorada entre elas, dificultando a sua classificação..



Outro fato que corrobora o problema de localização geral de qualquer ponto em qualquer posição da foto, são os gráficos de distribuição vertical e horizontal dos pontos a seguir. Neles podemos ver que horizontalmente existe uma distribuição uniforme em para todos os pontos marcados do corpo e verticalmente temos uma leve diferença.



Vemos uma outra questão nas imagens retiradas de cada classe a seguir. Nela, vemos ambientes, pessoas, iluminações e posições bastantes diferentes. Além disso, pessoas em movimentos como é o caso da imagem da classe *sports*, tendem a ter a imagem mais desfocada. Tudo isso são fatores que irão dificultar a busca de um padrão de aprendizagem a ser aplicado.



Um detalhe que pode ser confundido pela rede futuramente é que temos uma bicicleta na classe *miscellaneous*, que poderia ser aplicada em *sports*, facilmente. Esses são alguns problemas que temos que enfrentar para chegar ao objetivo de classificação e de localização das *keypoints*. Lembrando que algumas dessas características podem ser atenuadas ou fortalecidas com os filtros de pré-processamento ou de normalização que iremos aplicar na próxima etapa do projeto.