Avaliação do Modelo

Acerca do primeiro modelo

Este modelo treinou durante 50 épocas, e foi adicionado um callback para efetuar paragem de treino sempre que não houvesse melhoramento da “val\_accuracy” durante 10 épocas.

Com a observação dos gráficos, obtemos umas curvas não muito suaves, com alguns picos nos dados, no entanto percebemos que tanto a validação como o treino se acompanham lado a lado.

Quanto a loss e ao val\_loss, continuam sempre a cair, ao longo das 50 épocas, até valores perto de zero, no entanto, a loss, perto da época 50, começa a estabilizar enquanto que a val\_loss continua a descer depois de um pico.

Na accuracy, como era de esperar pois demostra o contrário da loss, a curva também não é linear, tem muitos altos e baixos, muitos picos, no entanto ela vai aumentando até ao valor 1, durante as 50 épocas. Se não fosse pelos altos e baixos das curvas, poderíamos estar perante um resultado muito bom, ou então um resultado enganoso, pois o modelo poderia ter “decorado os dados de treino.

No gráfico da mse, obtemos um comportamento e um gráfico muitíssimo semelhante ao obtido na los, sem a parte da estabilizar a mse perto da época 50.

Conseguimos obter uma avaliação nos dados de teste de 0.050 de loss, 0.0058 de mse e 1.0 de accuracy. No entanto, ao fazermos mais testes obtemos uma binária accuracy de 0.93 e uma precissision de 0.9.

Quanto ao teste para classificar uma nova imagem, em que carregamos essa imagem, obtemos a resposta pretendida, carregamos uma imagem com a rua limpa e o modelo classificou-a como clean.

No teste com as imagens em gray Scale, obtemos gráficos diferentes, nomeadamente a loss que, apesar de ser linear, sem os picos, cai acentuadamente nas primeiras iterações mas mantem-se a diminuir, muito suavemente, ao longo das35 iterações.

Na accuracy obtemos um gráfico com o mesmo formato do teste anterior, contendo muitos valores inconstantes nos dados de validação.

Quanto a mse, começa a cair drasticamente nas primeiras iterações, e durante as restantes epocas, continua a diminuir, mas de forma mais suave a par da mse\_val.

Neste teste obtemos uma accuracy de 0.87 e uma perda e 0.38, ou seja, valores piores do que os obtidos no teste anterior.

Obteve-se uma Binary accuracy de 0.73 e uma precision de 0.66, o que também é inferior aos resultados obtidos no modelo anterior.

Optou-se por se fazer um teste com imagens em tom cinza, apenas um canal de cor, pois o modelo poderia ter mais facilidade em lidar com os dados de entrada e obter melhores resultados, no entanto, pelos números observados, não foi o resultado esperado.

Por último, foi feita uma abordagem e efetuado um teste, os resultados não foram o esperados no entanto podemos usar este teste “quase falhado” para perceber que até uma boa abordagem necessita de ser bem implementada com os parâmetros certos.

Utilizamos o “data augmentation”, para “gerar” imagens no dataset de treino e assim permitir “aumentar” o nosso dataset. O resultado ficou muito abaixo do que esperado, como podemos observar pelos gráficos obtidos, onde obtivemos umas curvas não muito fáceis de interpretar e não muito normais.

Este teste treinou por 13 épocas, o callback foi chamada e terminou o treino, e obteve uma Binary accuracy de 0.53 e uma precission de 0.57.

Podemos concluir que, apesar de a abordagem ser boa e ser aplicável em casos em que temos poucas imagens de treino, não obteve melhores resultados do que o primeiro modelo.

Problemas na implementação e nos parâmetros selecionados para esta abordagem podem estar na causa deste mau resultado. Seria necessário mais tempo para analisar e testar melhor os parâmetros e as implementações desta abordagem.

- Com estes testes podemos concluir que poderemos melhorar o desempenho do modelo, utilizar outras técnicas para o tratamento das imagens ou também, algo que poderia ser muito importante neste caso, seria aumentar o dataset para contermos mais imagens e conseguirmos ter mais dados para treinar.

Podemos referir que:

- Poderíamos ter melhores resultados se o dataset contivesse um maior número de dados, pois neste momento apenas estávamos a trabalhar com pouco mais de 207 imagens.

- Experimentar novas métricas, novos parâmetros, fazer um ajuste muito mais intensivo aos parâmetros utilizados para o treino e na criação da camada convulacional.

- IUtilizar técnicas para visualização dos dados, para alem de gráficos, para ajudar a perceber onde o modelo possa estar a precisar de mais ajustes.

- Perceber melhor como implementar e como melhorar os parâmetros de técnicas que lidam com datasets com poucos dados para treino.

No entanto, todas estas abordagem não foram possíveis de colocar totalmente em prática, apenas de um modo mais suave, pois o modelo ainda precisa de algum tempo para processar os dados e fazer previsões. Para fazer um super ajuste de parâmetros iria ser necessário muito tempo.