Contagem de Peças

Bruno M. Pacheco

mpacheco.bruno@gmail.com

Introdução

Controle de estoque entre processos é um problema logístico presente em quase todas as indústrias. Na Metalúrgica FIMAC, que foca em usinagem de peças sob medida, um lote de uma dada geometria é armazenado em uma caixa de aço como na figura 1, utilizada para armazenar e transportar as peças entre processos. O problema de controle surge quando é necessário determinar a quantidade de peças em uma caixa, que, atual- mente, necessita da intervenção de um operador e se torna um processo custoso. Dessa forma, surgiu a proposta de analisar a viabilidade de utilizar métodos utilizando visão computacional para estimar a quantidade de peças em uma dada caixa.

Figura 1: Caixa com peças, objeto deste trabalho.

Métodos

A abordagem acordada foi de simular uma câmera que capturasse também a profundidade para cada pixel. A câmera seria posicionada acima da caixa, capturando a superfície superior formada pelas peças na caixa. Além disso, como a maior dificuldade surge quando há a oclusão completa de muitas peças, o problema foi formulado como uma tarefa de regressão a partir da imagem capturada ao invés de uma tarefa de segmentação ou detecção de objetos.

Dados

A partir da geometria da caixa e de uma peça, o posicionamento de múltiplas peças em uma caixa foi simulado e, posteriormente, renderizado a partir de uma câmera RGBD ideal. O Blender foi utilizado para ambas as tarefas, gerando um conjunto de 1000 simulações, em que 100 peças eram adicionadas na caixa em cada simulação. Dessa forma, o conjunto total possui 101000 imagens, uma vez que foi renderizada também uma imagem antes de ser adicionada qualquer peça. Um exemplo pode ser visto na figura 2. 20% das simulações foram reservadas para realizar os testes e avaliações finais, não sendo utilizadas em nenhum momento durante o desenvolvimento.

Figura 2: Uma das imagens do conjunto gerado. À esquerda, a imagem em tons de cinza. À direita, o canal de profundidade.

Rede Neural Convolucional

Uma das abordagens foi a de treinar uma rede neural convolucional baseada na EfficientNet via *transfer learning* seguido de *fine-tuning*. O encoder da EfficientNet foi mantido, enquanto o decoder foi substituído por uma camada completamente conectada seguida de um único nó, para a tarefa de regressão. Foram feitos experimentos com duas mudanças arquiteturais: tamanho da EfficientNet (b0 e b4) e tamanho da camada completamente conectada. Além disso, para cada arquitetura, foram feitos experimentos para otimizar a taxa de aprendizagem e seu decaimento, e a função de perda.

Visão Computacional Clássica

O primeiro passo para conseguir extrair as informações relevantes das peças no interior da caixa é conseguir segmentá-las. Dessa forma, uma abordagem para segmentar somente o interior da caixa levando em consideração seu aspecto tridimensional foi elaborado. Em suma, o processo segue:

1. Janelamento do canal de profundidade na região de interesse (na altura da borda da caixa)
2. Canny para extrair as bordas
3. Transformada de Hough para encontrar as linhas que demarcam a borda
4. A partir das linhas, definição do volume (paralelepípedo) que delimita o interior da caixa
5. Transformação da imagem RGBD em nuvem de pontos e recorte utilizando o volume calculado

Após essa etapa, têm-se não a novem de pontos gerada a partir da imagem RGBD original, mas já a segmentação somente da superfície definida pelas peças. A figura 3 ilustra essa diferença. A partir dessa superfície, duas abordagens foram tomadas para estimar a quantidade de peças: "cavar"a geometria da peça da superfície; estimar o volume ocupado e aplicar um método de regressão.

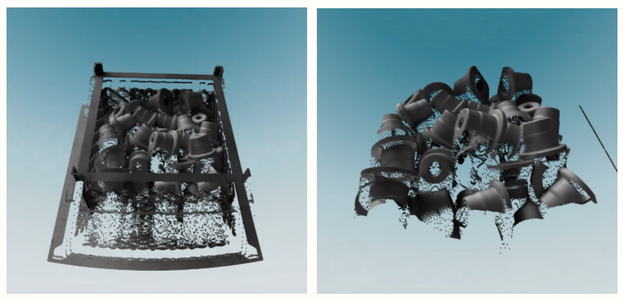

Figura 3: Nuvem de pontos da imagem RGBD. À esqueda, nuvem de pontos gerada a partir da imagem RGBD original. À direita, mesma nuvem de pontos após segmentação do interior da caixa.

Para a primeira abordagem, a premissa é de que podemos alinhar a geometria da peça de interesse (conhecida de antemão) a uma das peças na superfície encontrada e realizar a subtração, efetivamente "cavando"a superfície. Pode-se realizar esse processo repetidas vezes até que não haja mais volume para ser subtraído. A quantidade de peças subtraídas é a estimativa do número de peças.
De contra-partida, estimar o número de peças a partir do volume oferece uma solução mais simples já que diferentes arranjos de uma mesma quantidade de peças geram volumes percebidos diferentes. A partir do volume delimitado pela superfície encontrada, tanto uma regressão linear quanto uma regressão polinomial de segunda ordem foram aplicadas.

Experimentos e Resultados

Os diversos experimentos com a rede convolucional implementada estão disponíveis na plataforma Weights and Biases: transfer learning e fine-tuning. De forma breve, a arquitetura com maior desempenho utiliza a EfficientNet b0, com 60 nós na camada oculta do *decoder*. Ela foi treinada por 47 épocas com os pesos da EfficientNet congelados (*transfer learning*) e, após, mais 81 épocas (*fine-tuning*), totalizando 160 horas de treinamento.

Para a abordagem de alinhamento e subtração da geometria, foi testado o alinhamento através de algoritmos tradicionais de co-registro (RANSAC e ICP), *matching* do histograma de gradientes, e otimização por enxame de partículas (PSO). O alinhamento utilizando PSO foi o que apresentou melhores resultados, ainda que seja bastante custoso computacionalmente. Entretanto, como todas as abordagens são custosas, somente a com PSO foi avaliada no conjunto de testes. Já para a abordagem mais simples, de estimativa a partir do volume, tanto uma regressão linear quanto uma regressão polinomial de segunda ordem foram testadas.

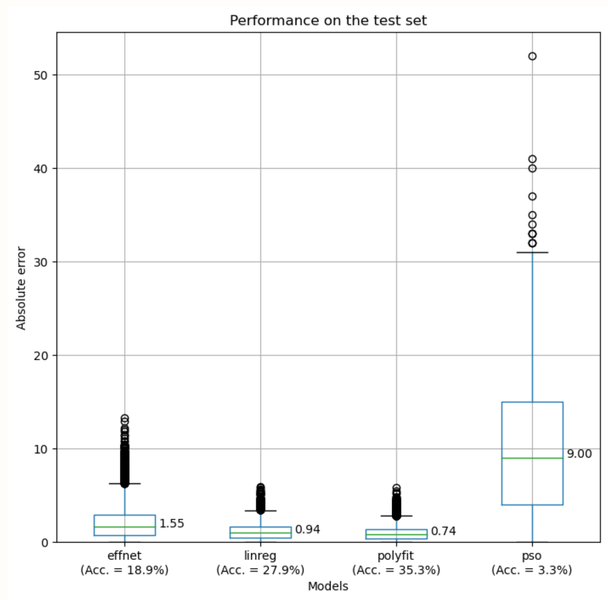

Figura 4: Performance dos modelos no conjunto de teste. Devido ao alto tempo de inferência, a abordagem utilizando PSO foi avaliada em um subconjunto aleatório do conjunto de teste.

As abordagens foram avaliadas quanto ao erro absoluto médio de suas predições. Como a estimativa é um número inteiro, a métrica foi calculada considerando o arredondamento das estimativas quando necessário. O desempenho das diferentes abordagens no conjunto de teste pode ser observado na figura 4.

Discussão

A abordagem utilizando aprendizagem profunda, ainda que tenha apresentado um desempenho de validação superior, apresentou uma queda significativa no conjunto de teste: um sinal de *overfitting*. Isso dá indícios que o uso de *data augmentation* pode ser bastante benéfico. Já a abordagem de alinhamento e subtração se mostrou inferior às demais. Sendo a abordagem mais custosa (levando cerca de 9 minutos por imagem), ela se torna inviável quando resoluções maiores (que trariam mais precisão à abordagem) são utilizadas. Ainda que exista margem para aprimoramento dessa abordagem, como o uso de algoritmos de co-registro para fazer um alinhamento prévio à PSO, dificilmente seria possível atingir resultados satisfatórios em um tempo de inferência adequado a uma aplicação real.
Finalmente, a partir dos resultados, fica evidente que nenhuma das abordagens mais complexas superou a estimativa a partir do volume encontrado, sendo a regressão polinomial a que melhor representou a relação entre o volume e o número de peças. Como a estimativa através do volume possui um erro inerente derivado do coeficiente de impactamento das peças na caixa, imagina-se que mesmo com uma melhor estimativa do volume, essas técnicas não têm margem para atingir um nível de acurácia suficiente para uma aplicação real. Portanto, sabendo do desempenho das demais abordagens, imagina-se que somente uma solução disruptiva conseguiria atingir o desempenho necessário.

Código: codigos.ufsc.br/bruno.m.pacheco/part-counting & Apresentação: youtu.be/g1NxVkaADig