



Método para Extração de *Background* e Estimativa de Pose 3D Baseado em Imagens de Profundidade

Autor: Matheus Alan Bergmann. Orientador: Claudio Rosito Jung

Universidade Federal do Rio Grande do Sul - UFRGS

Objetivos

- Estimar a pose da câmera em relação a um objeto conhecido com o uso de redes neurais convolucionais;
- Reduzir o erro causado pelo problema da circularidade;
- Reduzir o tempo necessário para treinar o modelo através de *Transfer Learning* e treino com *Loss Function* híbrida.

Visão Geral

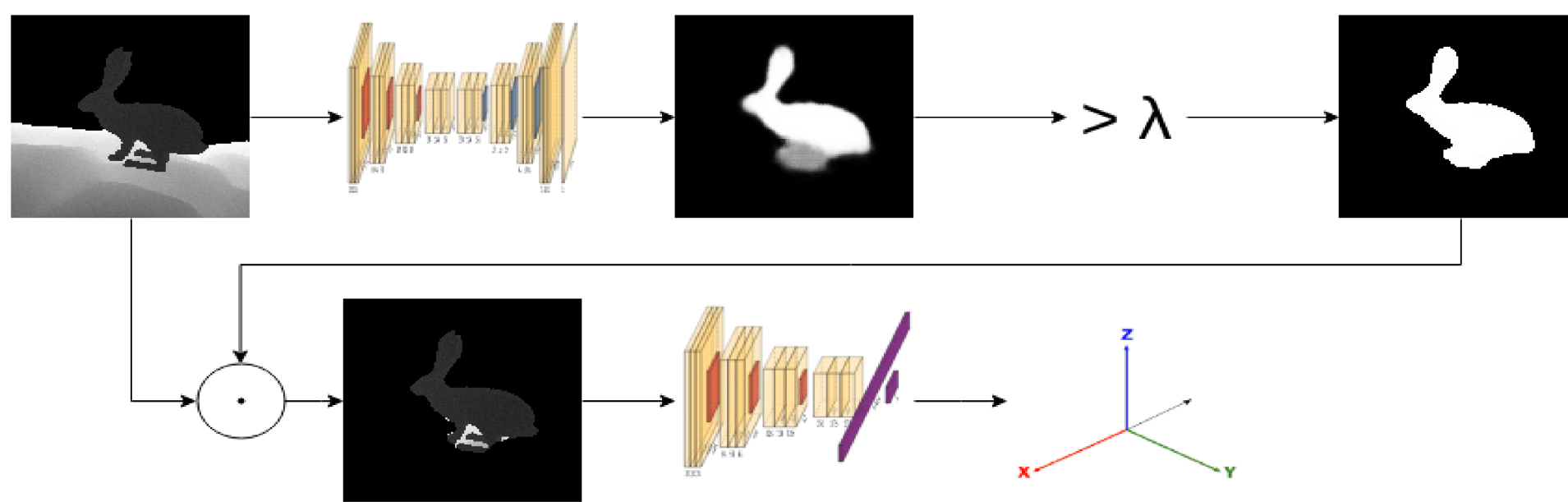


Figura 1: A entrada é o objeto em um fundo desconhecido. Ele passa por uma rede de segmentação que tem como saída a probabilidade de um pixel ser do objeto. Essa saída é binarizada e gera uma máscara de segmentação. Então a entrada é multiplicada ponto a ponto por essa máscara e o resultado passa pelo modelo de estimativa de pose que produz um vetor unitário que descreve a rotação.

A Estratégia Proposta

- A pose é representada como um vetor unitário que aponta do centro do objeto para a câmera;
- São estimados apenas dois graus de liberdade (2-DoF) visto que a inclinação longitudinal da câmera não é considerada;
- As vistas sintéticas do objeto são geradas posicionando a câmera em n pontos sobre um hemisfério centrado no objeto, como mostrado na Figura 2;
- Foi proposta uma rede profunda que realiza duas tarefas: segmentação e estimativa da pose.

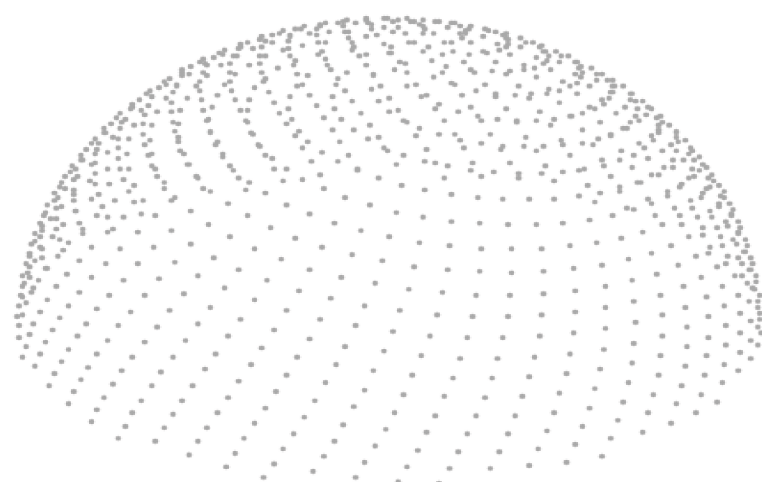


Figura 2: Posicionando a câmera sobre um hemisfério usando o Espiral de Fibonacci

Loss Function Híbrida

Nosso método é treinado com uma *Loss Function* \mathcal{L} parametrizada por um α que nos permite iniciar o treinamento da parte de segmentação e posteriormente refiná-lo enquanto a parte de estimativa de pose é treinada. CCE representa *Categorical Cross Entropy* e MSE o erro médio quadrado.

$$\mathcal{L} = \alpha CCE(y_{seg}, \hat{y}_{seg}) + (1 - \alpha) MSE(y_{pose}, \hat{y}_{pose}), \quad (1)$$

Experimentos

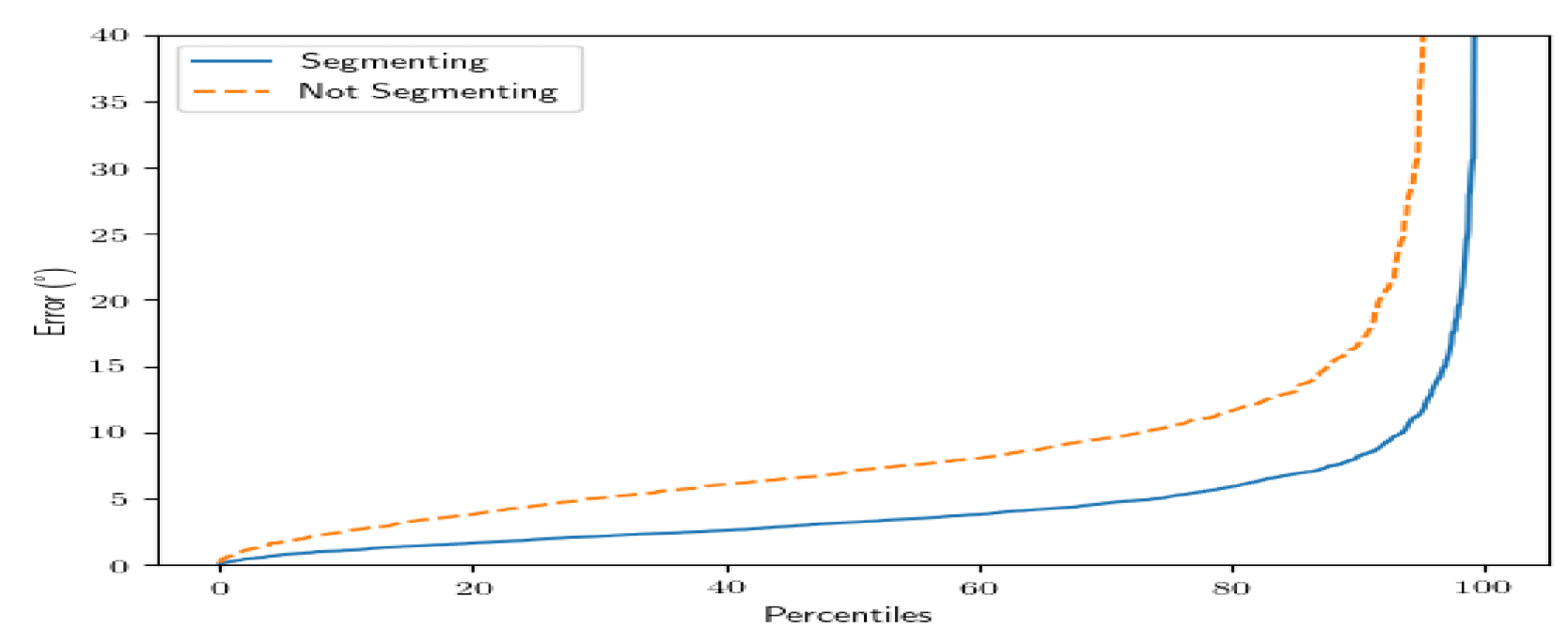


Figura 3: A imagem mostra como o erro varia para cada percentil quando a entrada do modelo foi segmentado (*background* removido) ou não.

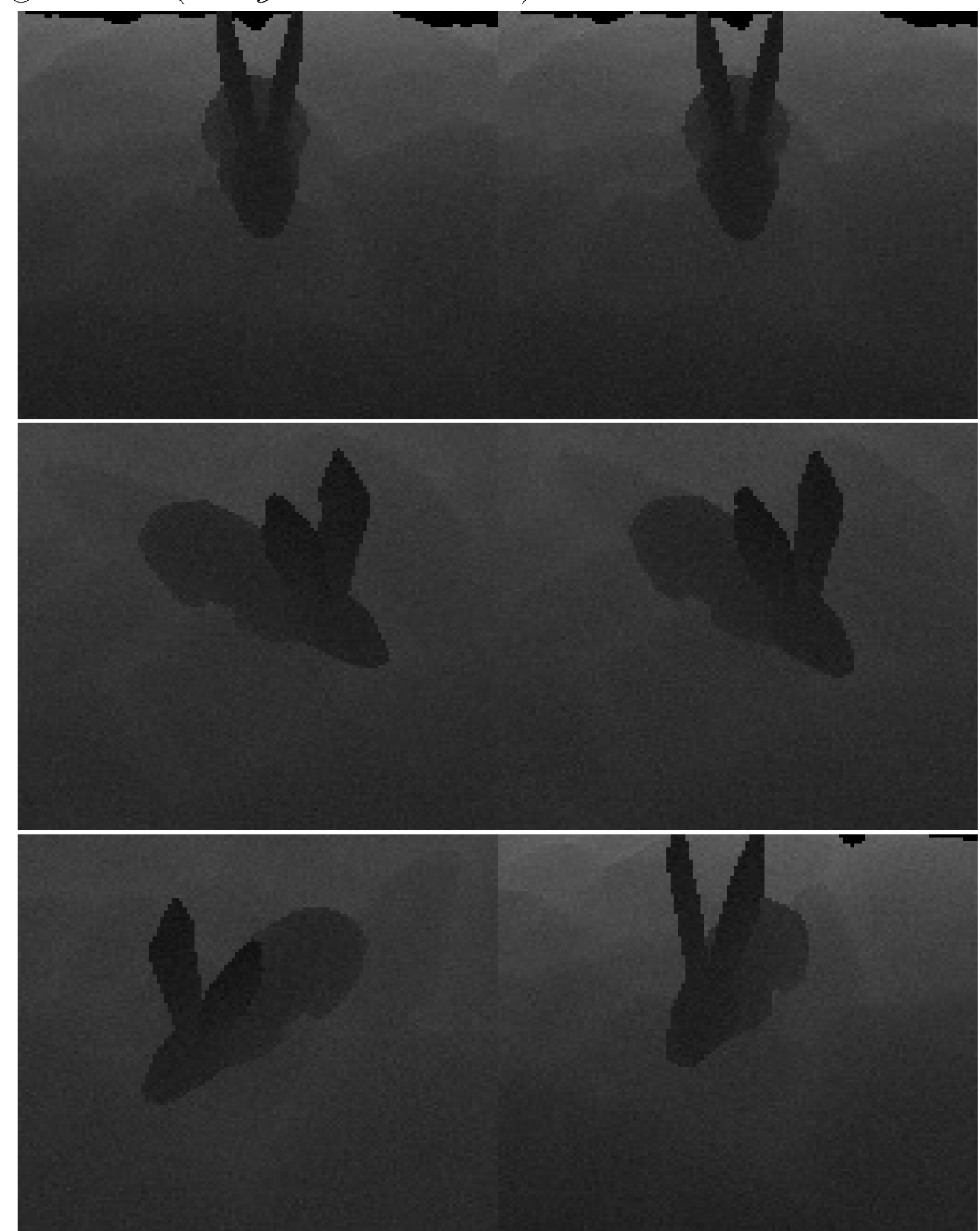


Figura 4: Pose original (esquerda) e pose estimada (direita). Os erros angulares são, de cima para baixo, 0,5452°, 3,2693°, 21,1209°