Using my own dataset:  
Epoch [1/10000], Train Loss: 5.1221, Train Accuracy: 1.56%

Validation Accuracy: 1.97%, Validation Loss: 5.0497

Epoch [2/10000], Train Loss: 4.9778, Train Accuracy: 2.26%

Validation Accuracy: 2.49%, Validation Loss: 4.9793

Epoch [3/10000], Train Loss: 4.9159, Train Accuracy: 2.89%

Validation Accuracy: 2.23%, Validation Loss: 5.0065

Epoch [4/10000], Train Loss: 4.8896, Train Accuracy: 3.02%

Validation Accuracy: 3.04%, Validation Loss: 4.9289

Epoch [5/10000], Train Loss: 4.8786, Train Accuracy: 3.20%

Validation Accuracy: 1.74%, Validation Loss: 5.0969

Epoch [6/10000], Train Loss: 4.8718, Train Accuracy: 3.21%

Validation Accuracy: 0.98%, Validation Loss: 6.2225

Using tiny imagenet:  
Epoch [1/10], Train Loss: 5.0684, Train Accuracy: 2.37% Validation Accuracy: 4.04%, Validation Loss: 4.7328

Epoch [2/10], Train Loss: 4.5978, Train Accuracy: 5.36% Validation Accuracy: 7.26%, Validation Loss: 4.4122 Epoch

[3/10], Train Loss: 4.3582, Train Accuracy: 7.84% Validation Accuracy: 8.98%, Validation Loss: 4.2329 Epoch

[4/10], Train Loss: 4.2156, Train Accuracy: 9.55% Validation Accuracy: 10.69%, Validation Loss: 4.1189 Epoch

[5/10], Train Loss: 4.0994, Train Accuracy: 10.80% Validation Accuracy: 12.12%, Validation Loss: 4.0186 Epoch

[6/10], Train Loss: 3.9795, Train Accuracy: 12.54% Validation Accuracy: 14.45%, Validation Loss: 3.8623 Epoch

[7/10], Train Loss: 3.8619, Train Accuracy: 14.28% Validation Accuracy: 15.11%, Validation Loss: 3.7965 Epoch

[8/10], Train Loss: 3.7658, Train Accuracy: 15.61% Validation Accuracy: 16.36%, Validation Loss: 3.7135  
  
  
  
using imagenet 5, and optimizer = optim.Adam(model.parameters(), lr=0.001)  # Adam optimizer  
Epoch [1/100], Train Loss: 4.0524, Train Accuracy: 13.76%

Validation Accuracy: 18.09%, Validation Loss: 3.6925

Epoch [2/100], Train Loss: 3.2865, Train Accuracy: 25.65%

Validation Accuracy: 27.61%, Validation Loss: 3.1561

Epoch [3/100], Train Loss: 2.8863, Train Accuracy: 32.88%

Validation Accuracy: 32.76%, Validation Loss: 2.8978

Epoch [4/100], Train Loss: 2.5930, Train Accuracy: 38.60%

Validation Accuracy: 37.28%, Validation Loss: 2.6849

Epoch [5/100], Train Loss: 2.3453, Train Accuracy: 43.46%

Validation Accuracy: 39.64%, Validation Loss: 2.5772

Epoch [6/100], Train Loss: 2.1154, Train Accuracy: 48.05%

Validation Accuracy: 41.82%, Validation Loss: 2.4692

Epoch [7/100], Train Loss: 1.8927, Train Accuracy: 52.74%

Validation Accuracy: 42.10%, Validation Loss: 2.4927

Epoch [8/100], Train Loss: 1.6746, Train Accuracy: 57.32%

Validation Accuracy: 43.34%, Validation Loss: 2.5009

Epoch [9/100], Train Loss: 1.4356, Train Accuracy: 62.50%

Validation Accuracy: 43.94%, Validation Loss: 2.5334

Epoch [10/100], Train Loss: 1.1993, Train Accuracy: 67.77%

Validation Accuracy: 43.40%, Validation Loss: 2.6803  
  
overfitting

using imagenet 5, and optimizer = optim.Adam(model.parameters(), lr=0.001)  # Adam optimizer, weight\_decay=1e-4)