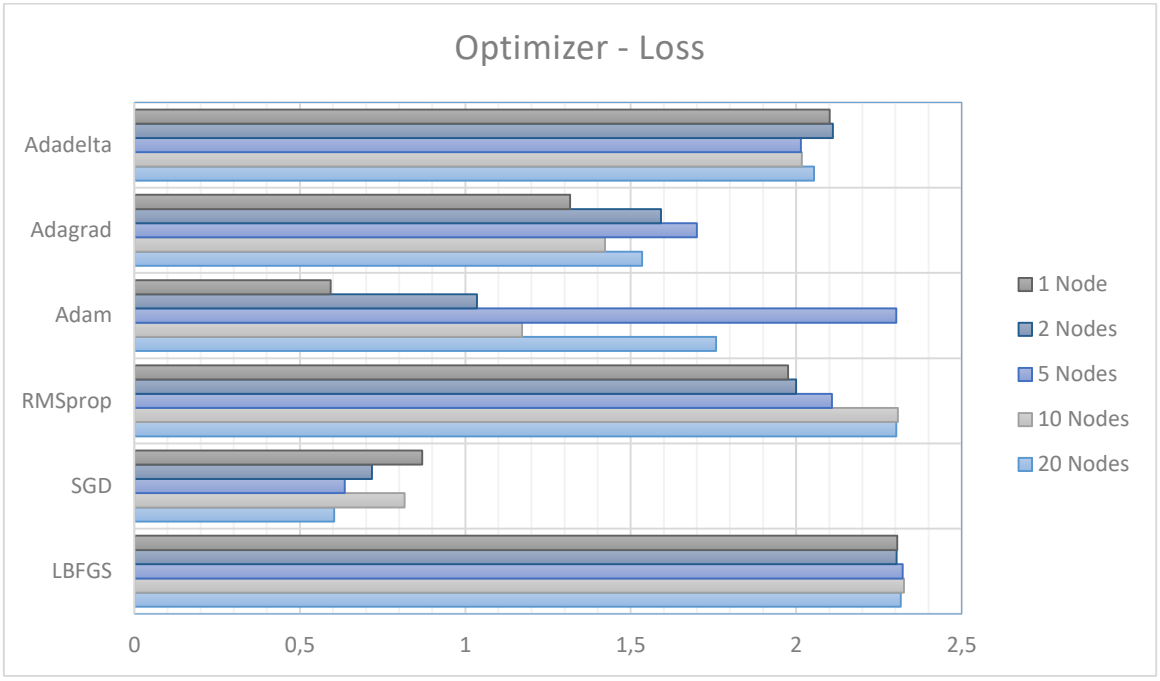
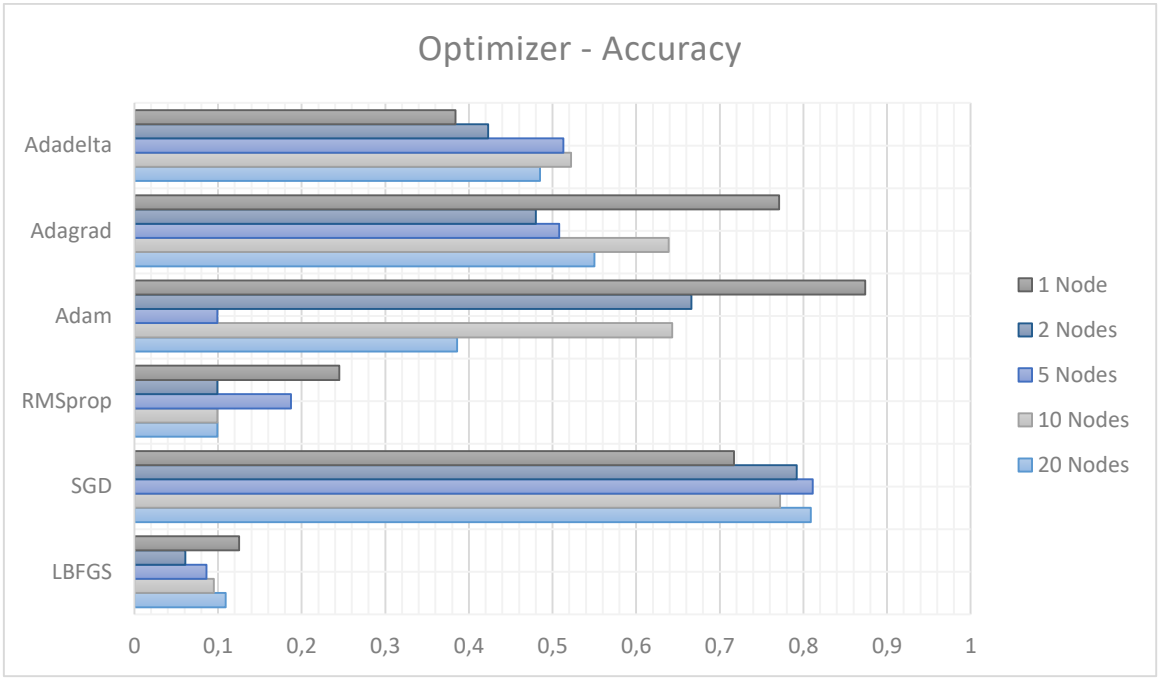
2D-Vision Aufgabe 4





Comparison of the optimizer

Adadelta	Adagrad	Adam	RMSprop	SGD	LBFGS
•More robust extension of Adagrad •Adapts learning rates based on a moving windows gradient updates, instead of accumulating all past gradients	•Adapts the learning rate to individual features •Some of the weights in the dataset will have a different learning rates •Learning rate tends to get really small over time	•Another way of using past gradients to calculate current gradients •Utilizes the concept of momentum by adding fractions of previous gradients to the current one	•Special version of Adagrad •It only accumulates gradients in a fixed window	•Implements stochastic gradient descent •Optionally with momentum	•Uses LBFGS-Algorithm which is in the family of quasi-Newton methos •Uses an estimation to the inverse Hessian matrix to steer its search through variable space