

Klasyfikacja cyfr na podstawie zdjęć Google Street View

Łukasz Neumann

Witold Oleszkiewicz

Dane wejściowe

Obrazki 32x32 z pojedynczą cyfrą



Wykorzystane środowiska

- Lasagne + Theano
- Keras + Theano/Tensorflow
- Dodatkowo numpy, scikit, matplotlib

Architektura sieci

Warstwa	Wymiary wyjścia	Liczba parametrów	Opis
conv 2D	30x30x32	320	32 maski 3x3
dropout	30x30x32		usunięcie 20% neuronów
conv 2D	28x28x32	9248	32 maski 3x3
max pool 2D	14x14x32		okno 2x2
conv 2D	12x12x32	9248	32 maski 3x3
dropout	12x12x32		usunięcie 20% neuronów
conv 2D	10x10x32	9248	32 maski 3x3
max pool 2D	5x5x32		okno 2x2
flatten	800		
dense	512	410112	
dropout	512		usunięcie 50% neuronów
dense	10	5130	

Architektura sieci

- Funkcja aktywacji warstw konwolucyjnych: ReLU
- Funkcja aktywacji warstwy wyjściowej: Softmax
- Algorytm stochastycznego spadku gradientu z momentum Nesterov'a
- Funkcja strat: entropia krzyżowa

Badanie dropout

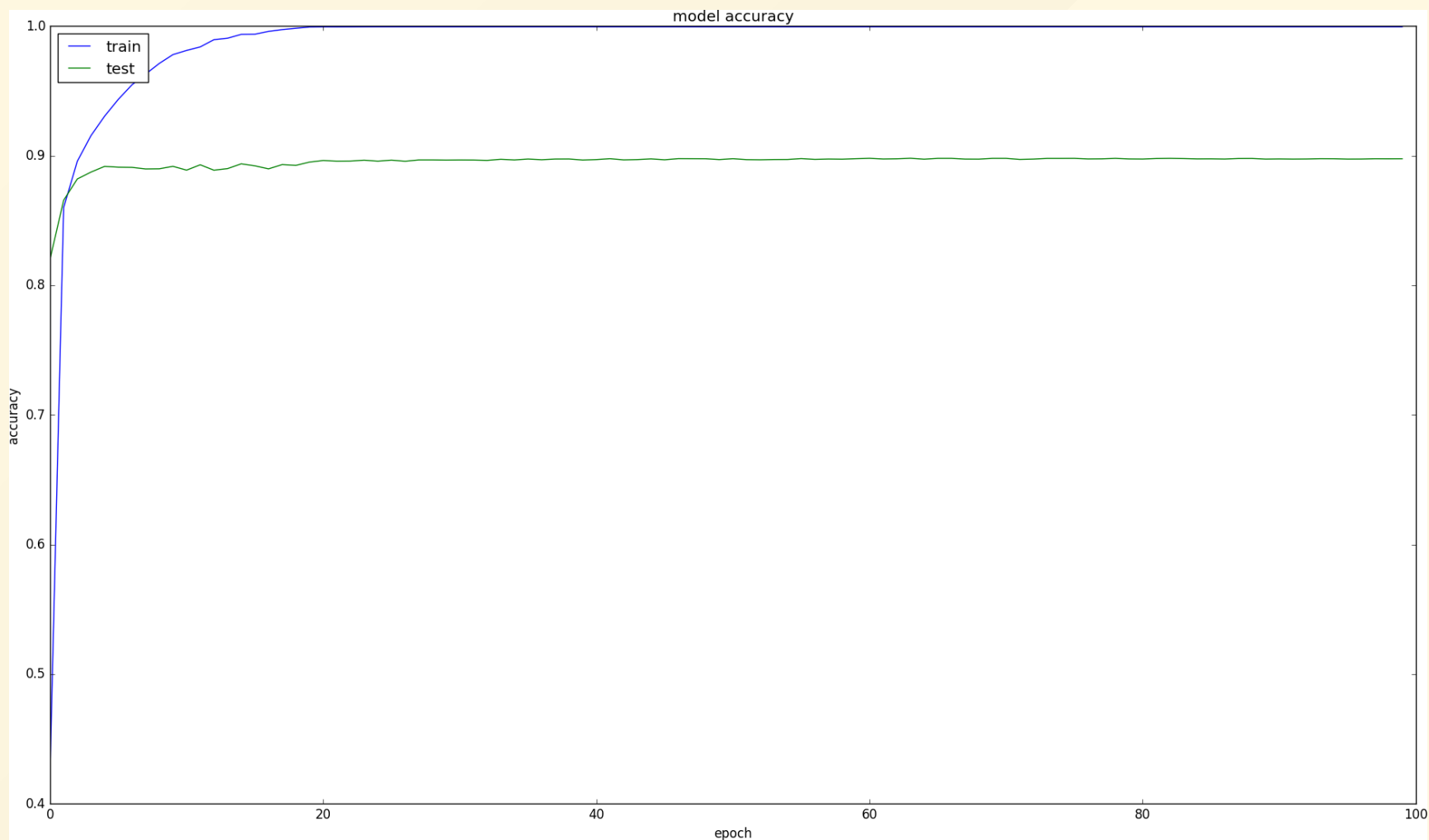
'Mały dropout

- 10% dropout na warstwach konwolucyjnych
- 25% dropout na warstwach

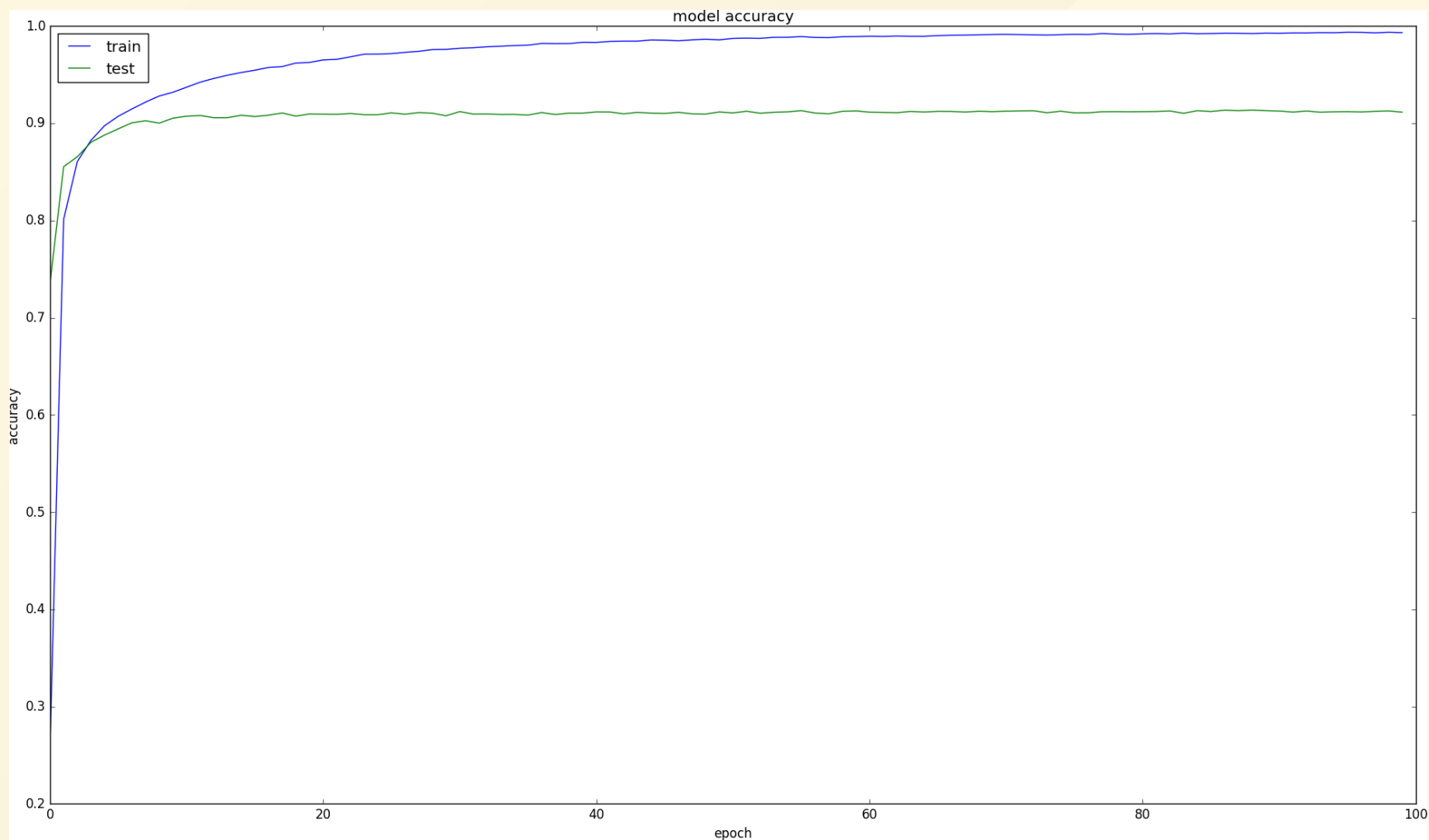
'Duży' dropout

- 25% dropout na warstwach konwolucyjnych
- 50% dropout na warstwach

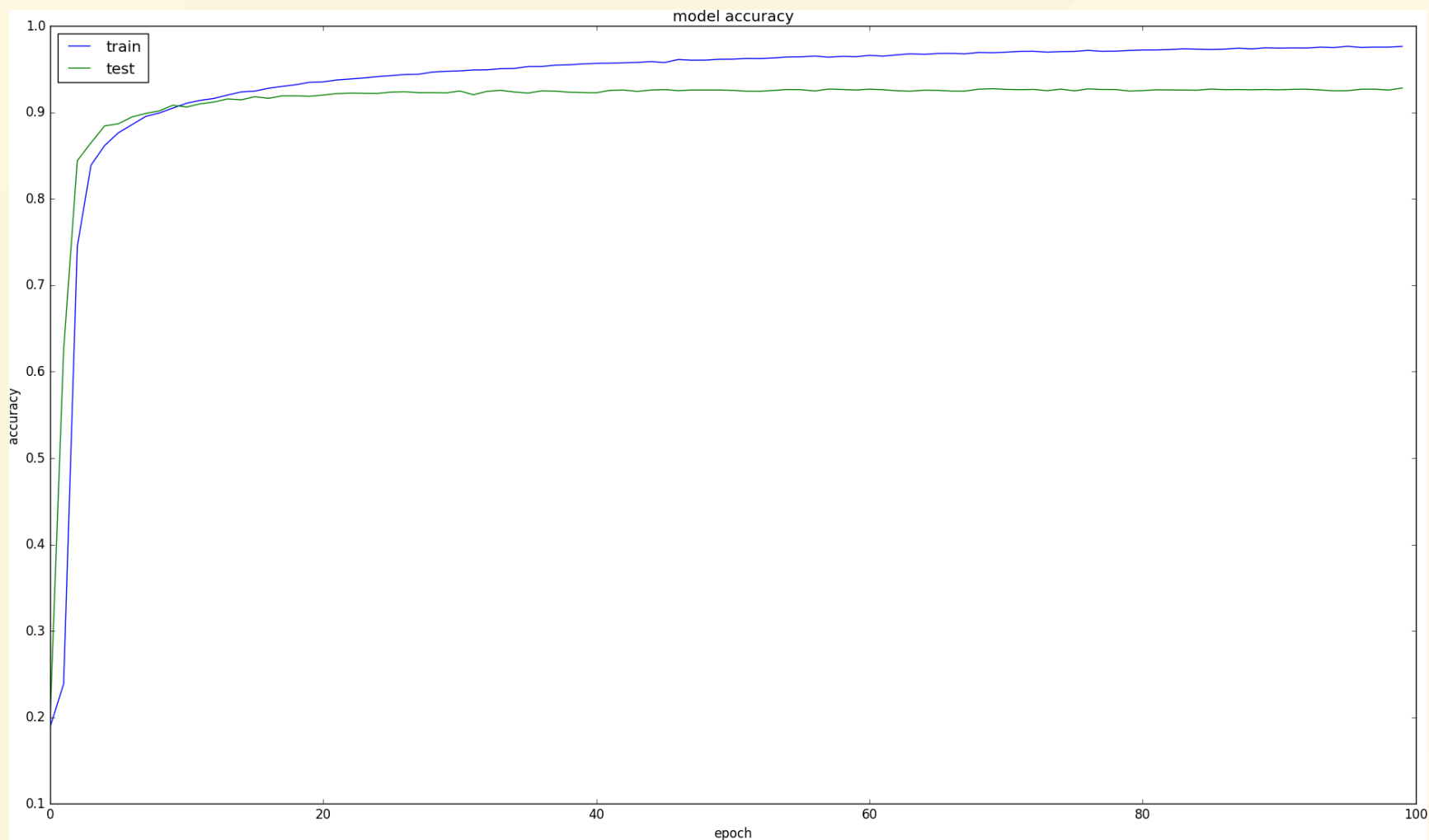
Wyniki celności bez dropout



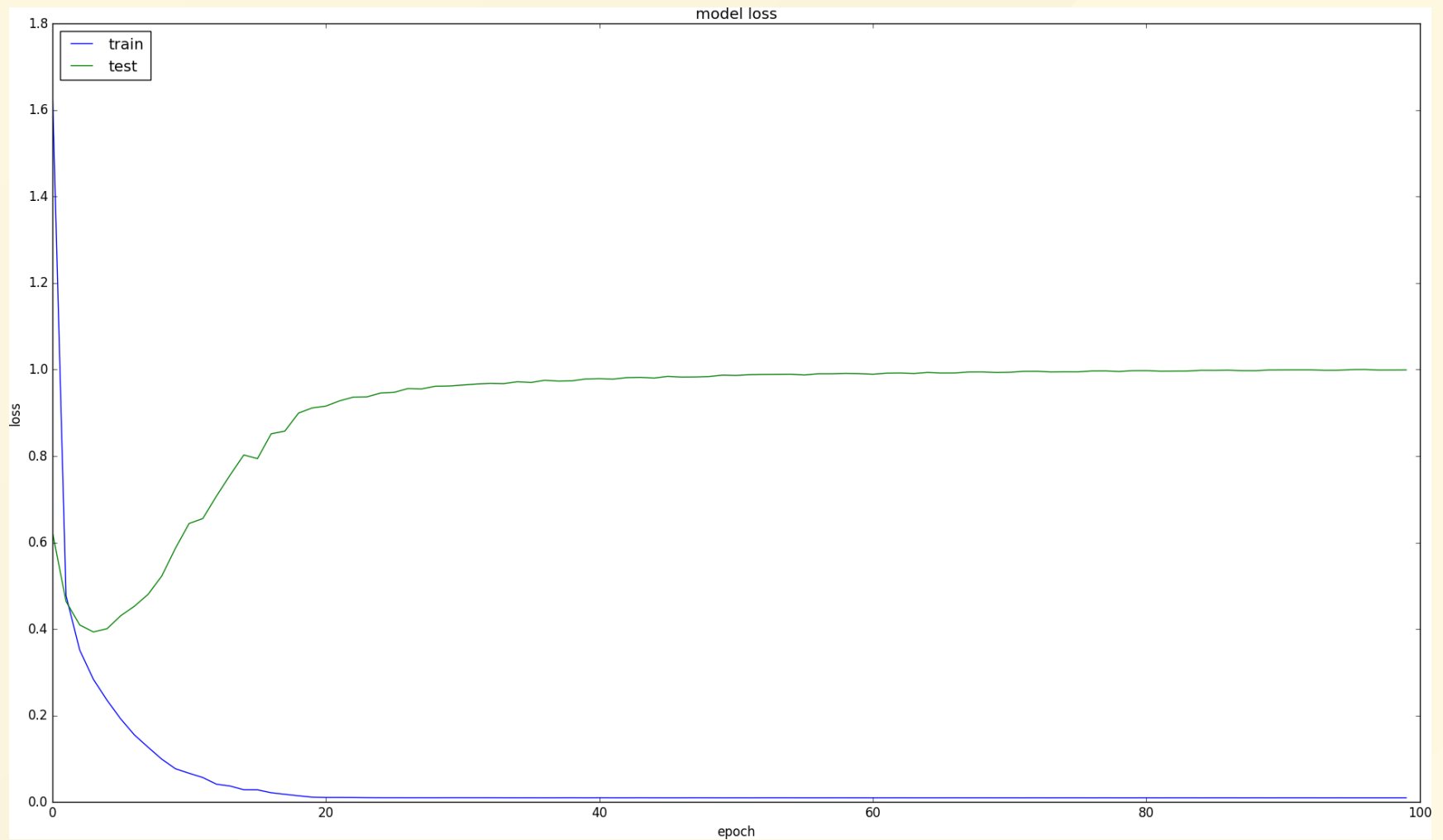
Wyniki celności - 'mały' dropout:



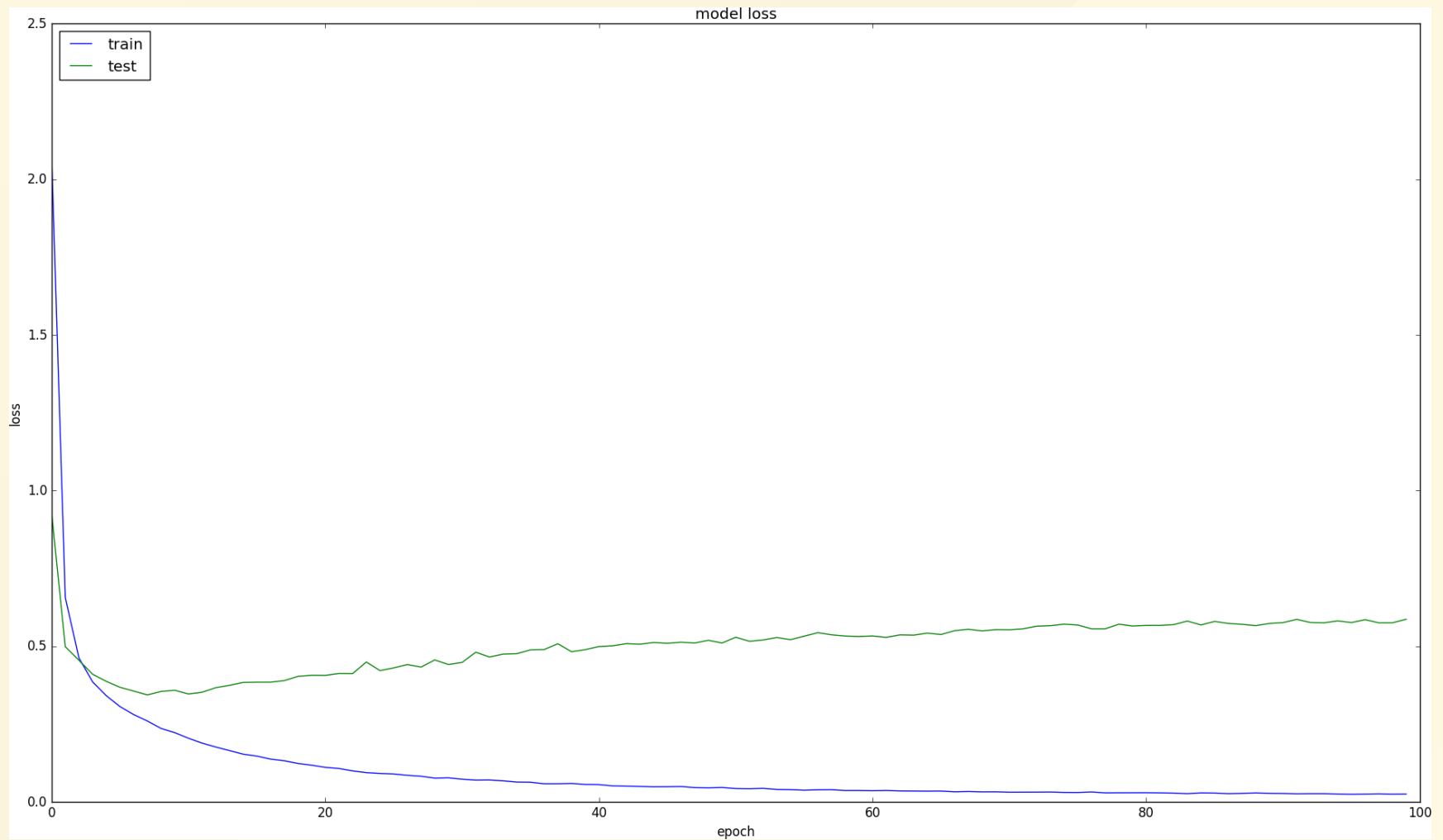
Wyniki celności - 'duży' dropout



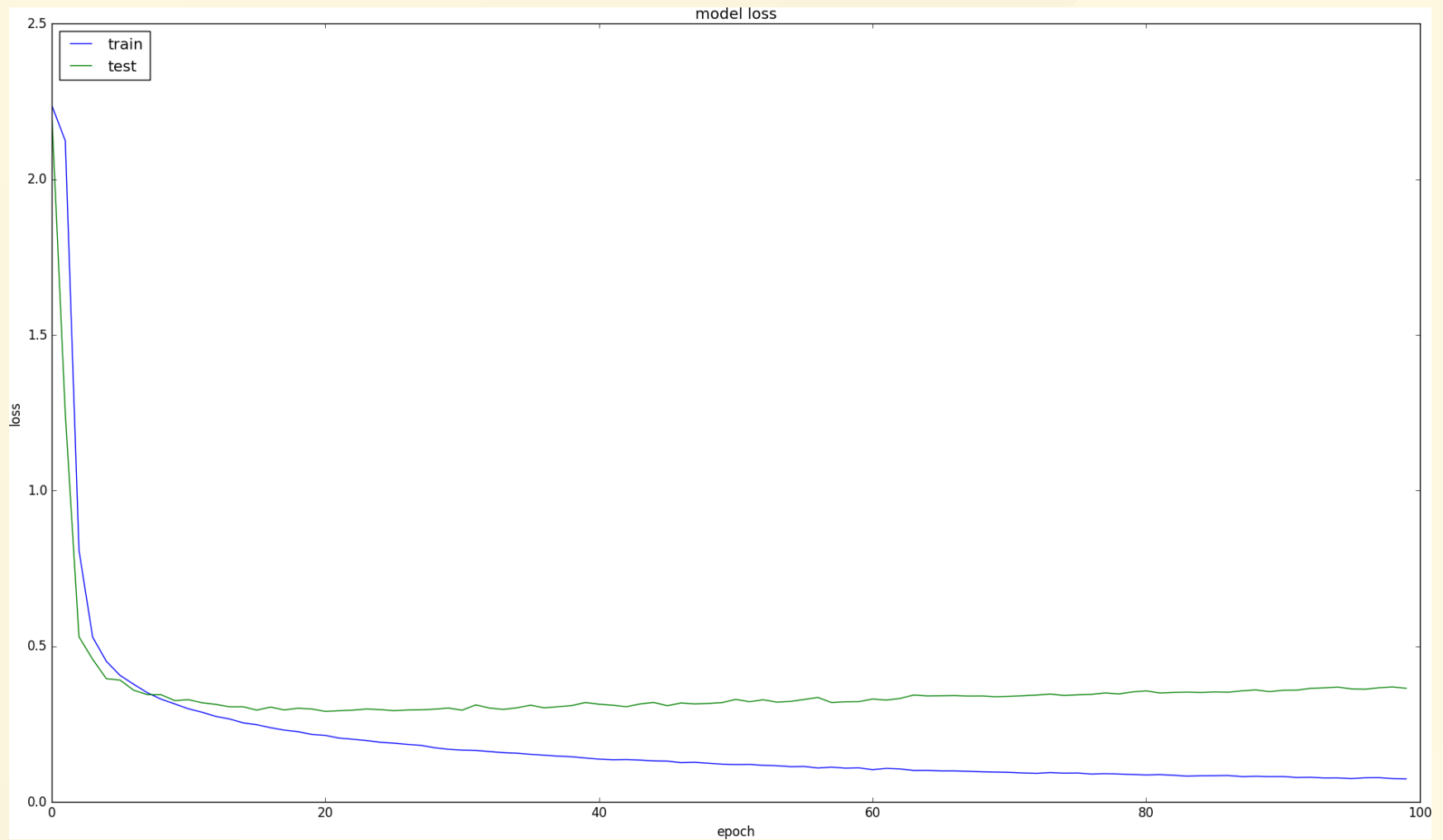
Funkcja strat bez dropout



Funkcja strat - 'mały' dropout



Funkcja strat - 'duży' dropout



Wyniki referencyjne w literaturze

- 93% celności dla sieci o podobnej liczbie warstw
- 96% celności dla większych sieci (17 warstw)

Ian J. Goodfellow, Yaroslav Bulatov, Julian Ibarz, Sacha Arnoud, Vinay D. Shet

"Multi-digit Number Recognition from Street View Imagery using Deep Convolutional Neural Networks", CoRR (2013)

Podsumowanie

- Dropout pozytywnie wpływa na generalizację modelu
- Wyniki porównywalne z rozwiązaniami referencyjnymi (~92.8% celności)
- Tensorflow jest szybszy na CPU
- Keras posiada wygodniejszy interfejs użytkownika w stosunku do Lasagne