

# Langevin Monte Carlo

- Initialize weights  $w^0$
- Do say 100 iterations with usual SGD, but add Gaussian noise  $\eta^k \sim \mathcal{N}(0, 2\varepsilon I)$  to each update
- After 100 hundred epochs decide that Markov Chain converged and start collecting weights values
- ~~For a new object predict compute average prediction of CNNs with weights  $w^{100}, w^{101}, \dots, w^{200}$~~
- Train another CNN to mimic the ensemble [Balan, Anoop Korattikara, et al. "Bayesian dark knowledge." *Advances in Neural Information Processing Systems*. 2015.]