Out-of-Class Novelty Generation: an Experimental Foundation

Mehdi Cherti & Balázs Kégl (LAL/LRI, CNRS/Université Paris Saclay)
Akın Kazakçı (MINES ParisTech, PSL Research University)

Research questions

- What is meant by the generation of novelty?
- How can novelty be generated?
- • How can a model generating novelty be evaluated?

Contributions

In Kazakçı et al. 2016:

- We show that symbols of **new types** can be generated by carefully tuned **autoencoders**
- We make a first step of defining a **conceptual and experimental framework** of novelty generation
- **However, we make no attempt to design evaluation metrics**

In this paper,

- We design an experimental framework based on **hold-out classes**
- We review and analyze the most common evaluation techniques from the point of view of measuring **"out-of-distribution novelty"** and **propose new ones**
- We run a large-scale experimentation to study the **capacity for generating novelty** of a wide set of generative models

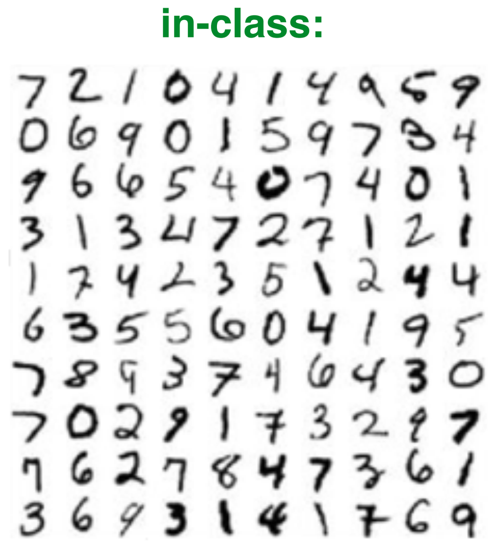
Setup

In our experiments:

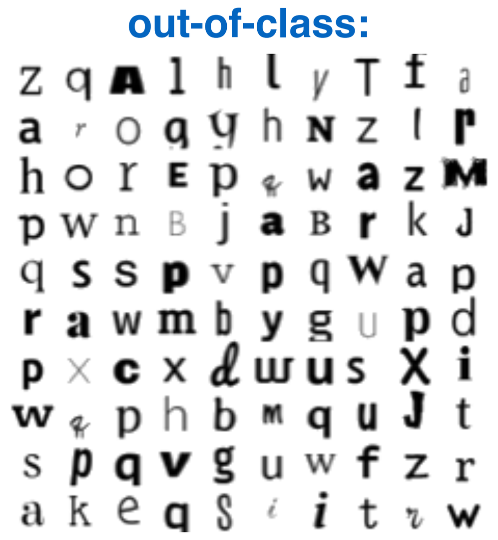
We train models on **digits**

We seek for models that generate **letters**

in-class:



out-of-class:



Evaluation metrics

Objectness (Salimans et al. 2016)

K : number of classes
n : number of examples

$$\frac{1}{N}\sum_{i=1}^{n}\sum_{\ell=1}^{K} p_{i,\ell}\log\frac{p_{i,\ell}}{p_\ell}$$

where:

$$p_{i,\ell} = p(l|x_i)$$

is the posterior probability of category l given the generated object x_i

and

$$p_l = \frac{1}{n}\sum_{i=1}^{n} p_{i,l}$$

are class marginals

Count and max

in-class posteriors:

$p_{i,1},\dots,p_{i,K_{\text{in}}} \longrightarrow$ 10 (digits)

out-of-class posteriors:

$p_{i,K_{\text{in}}+1},\dots,p_{i,K_{\text{in}}+K_{\text{out}}} \longrightarrow$ 26 (letters)

the most likely category overall:

$$l_i^* = argmax\ p_{i,l}$$

the most likely out-of-class category:

$$\tilde{p}_\ell = \frac{\sum_{i=1}^{n}\mathbb{I}\{\ell = \ell^*_{\text{out}\,i}\}}{\sum_{i=1}^{n}\mathbb{I}\{\ell^*_{\text{out}\,i} > K_{\text{in}}\}}$$

the diversity term:

$$-\frac{1}{\log K_{\text{out}}}\sum_{\ell=K_{\text{in}}}^{K_{\text{in}}+K_{\text{out}}}\tilde{p}_\ell\log\tilde{p}_\ell$$

Count:

$$(1-\lambda)\times\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}\{\ell_i^* > K_{\text{in}} \wedge p_{i,\ell_i^*} > \theta\} + \lambda\times\text{diversity}$$

Max :

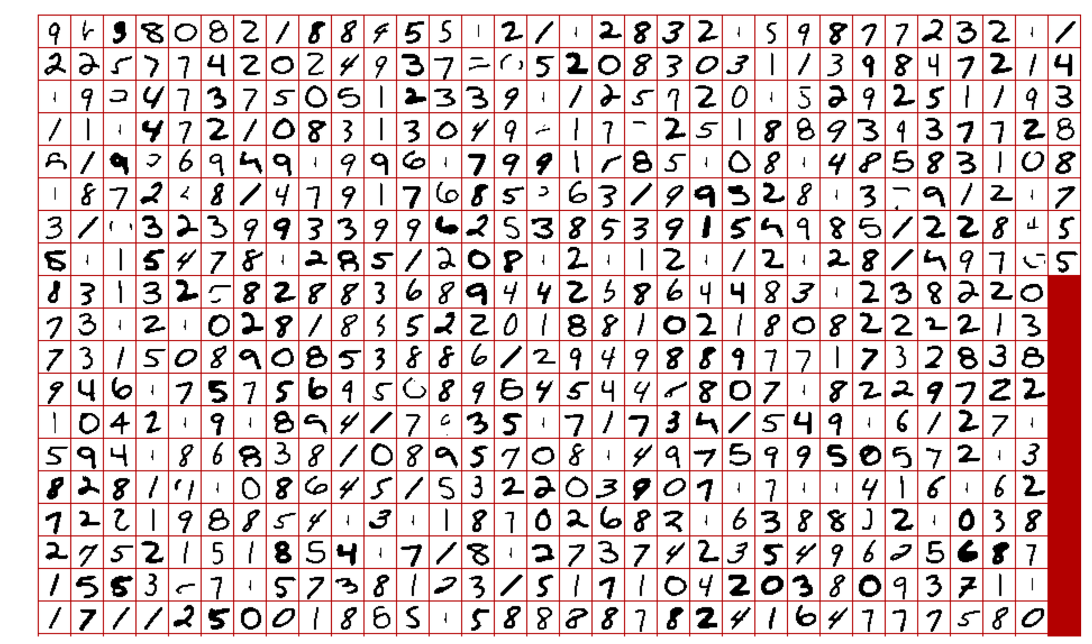
$$(1-\lambda)\times\frac{1}{n}\sum_{i=1}^{n} p_{i,\ell^*_{\text{out}\,i}} + \lambda\times\text{diversity.}$$

Experiments

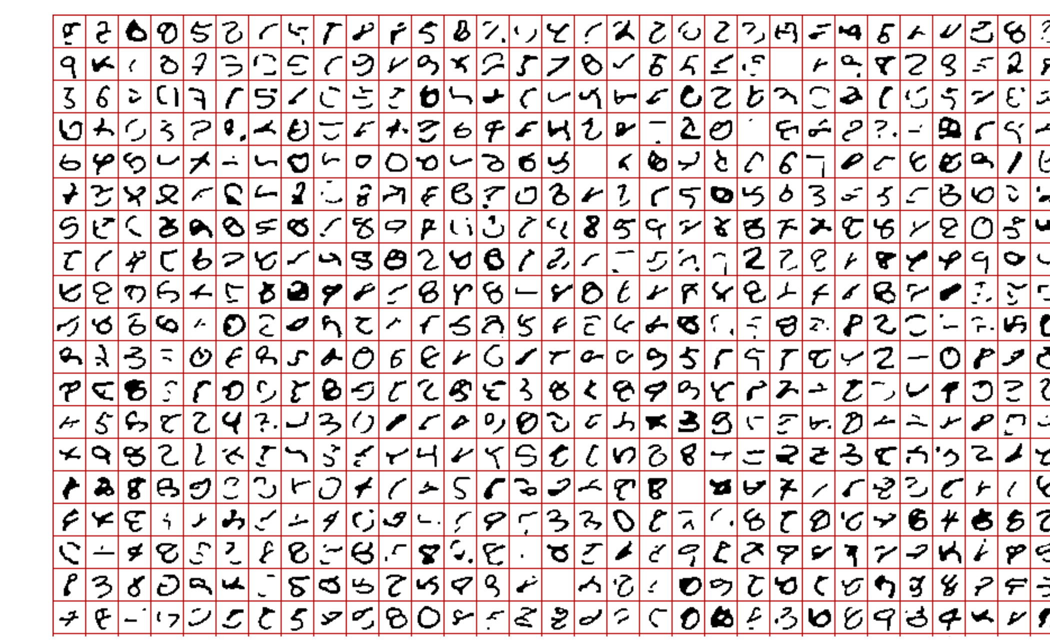
- We do a large scale experiment where we train ~1000 models by varying their parameters
- we use various kinds of **autoencoders** and **GANs**
- from each model, we generate 1000 images, then we evaluate the model using our proposed metrics
- We collect a total of ~1.000.000 generated images

Results

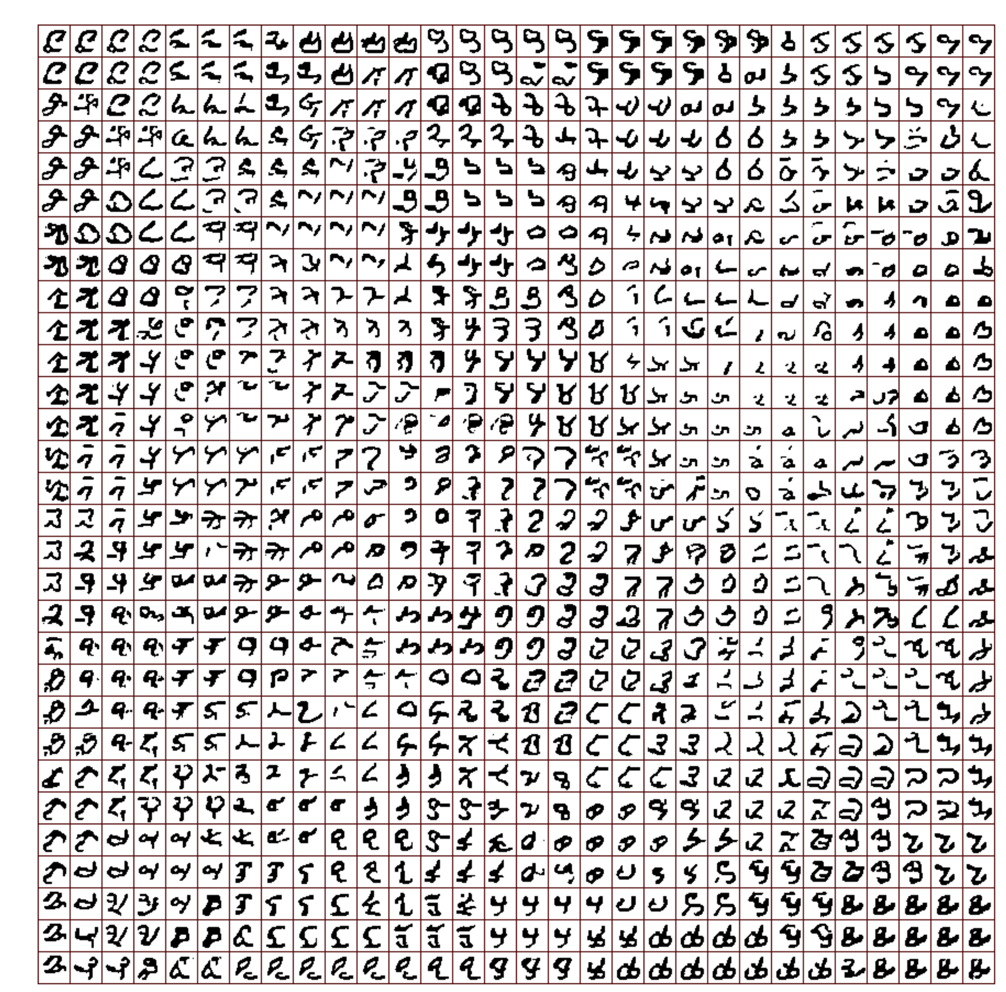
in-class selection

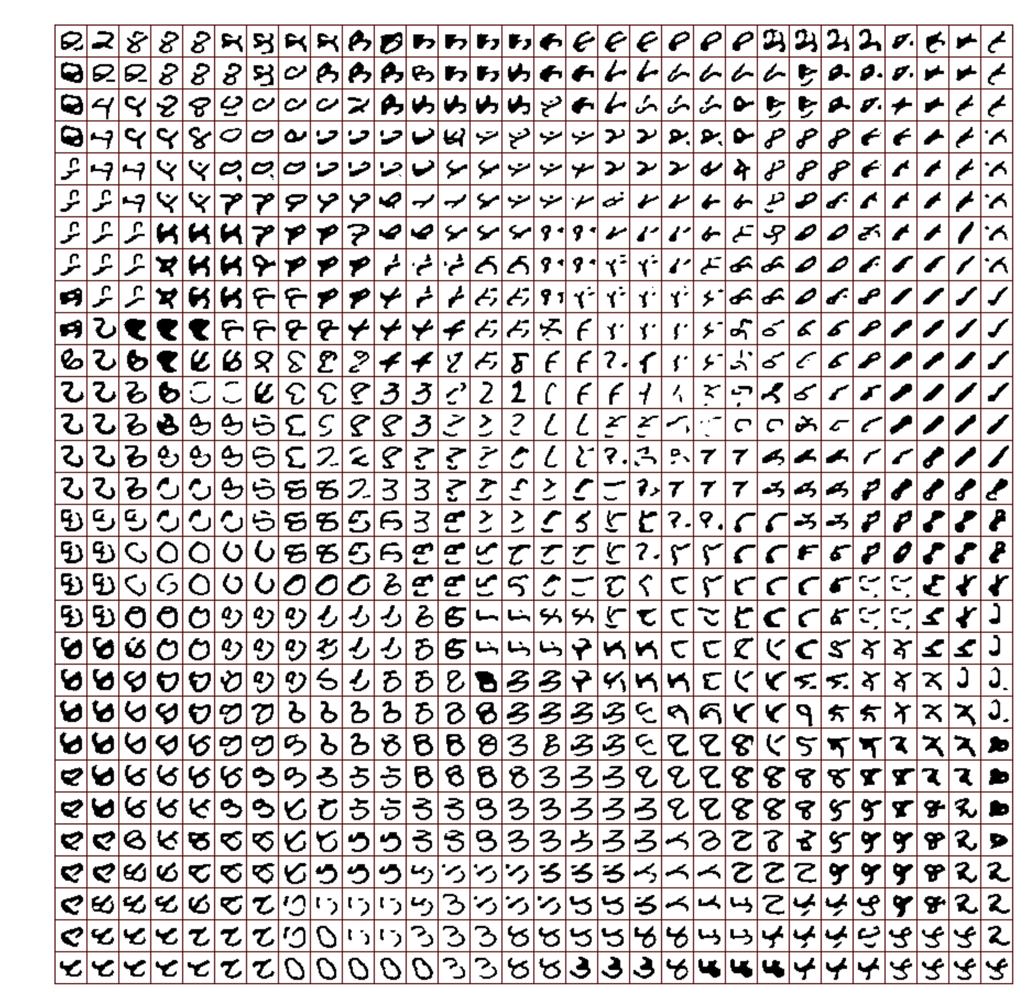


out-of-class selection

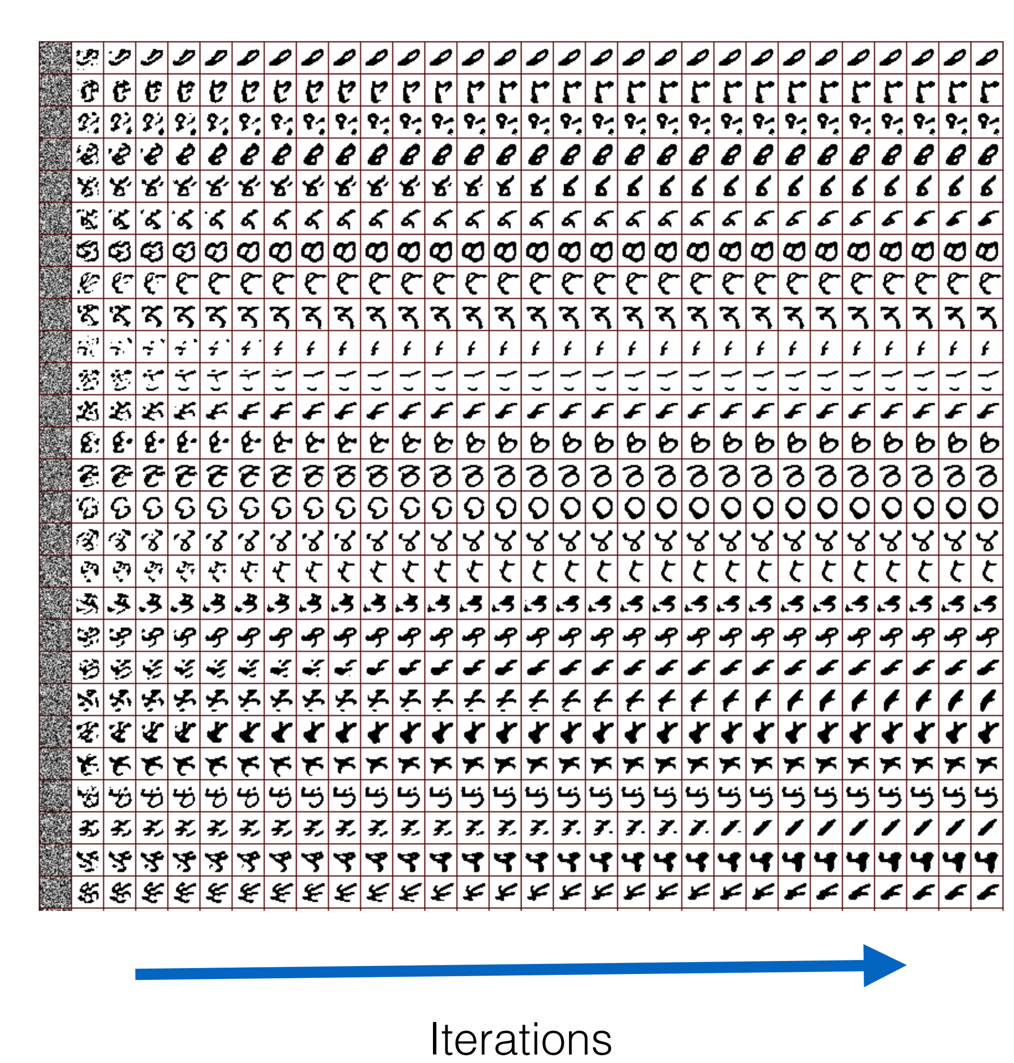


Visualization of samples





Generation



Letters