



Foundations models and MLOps

Maciej Durzewski

What a foundation model is?

- Big model trained on broad set of data.
- But does it tell us something?

Let's start from begin

- Everything has a purpose
- Everything generates problems
- Everything costs money

Typical problems with NN?

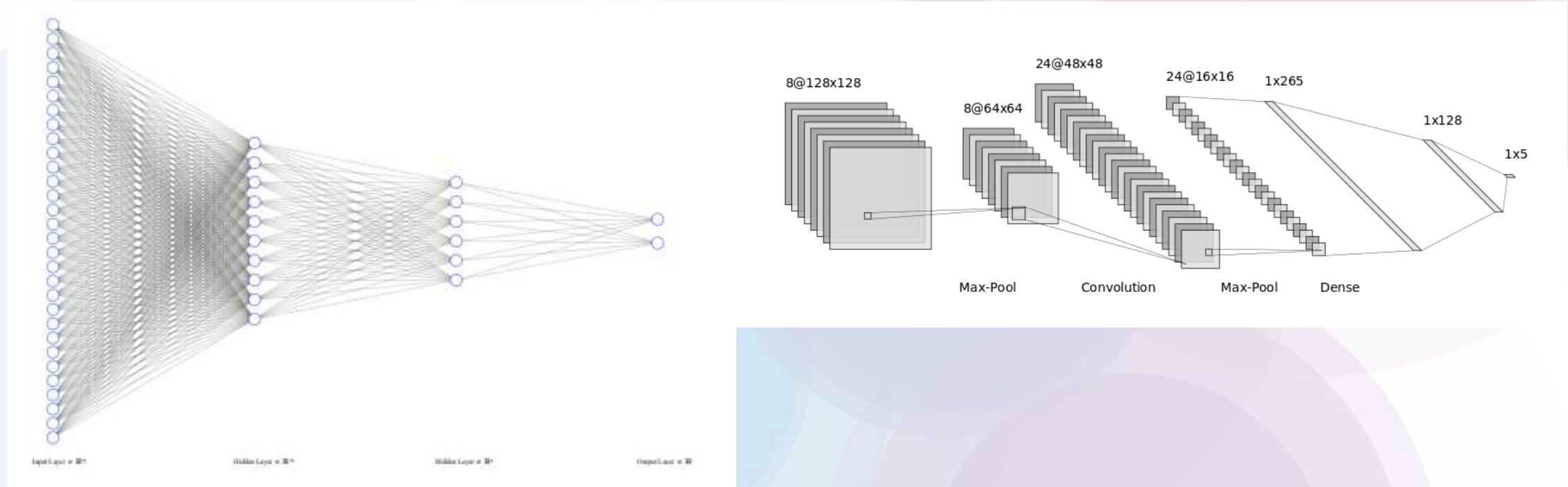
- Very sensitive on many parameters
 - architecture
 - functions
 - weight distributions
 - ... and many more
- Very sensitive because of data features
 - balance
 - order
 - parameters
 - ... and many more (not every data set is a mnist data set)
- The bigger model is the problems are bigger
- The bigger data sets are the problems are much bigger

New model new problems

- We need to prepare data again
- We need to adjust parameters again
- ...
- ...
- But at the end most of data sets are similar

How we can reduce amount of work?

- Let's take a look on how typical NN looks like



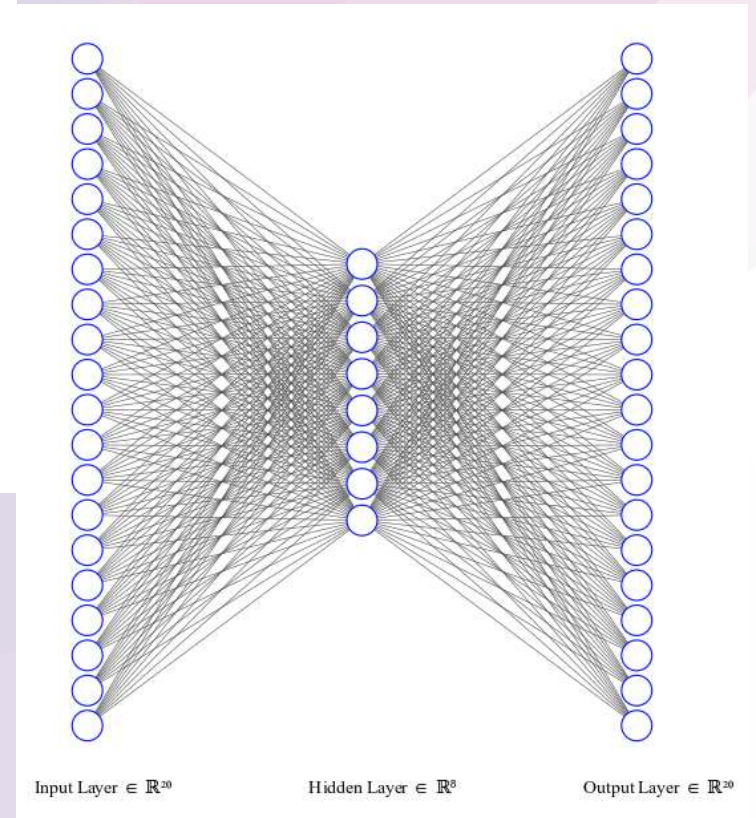
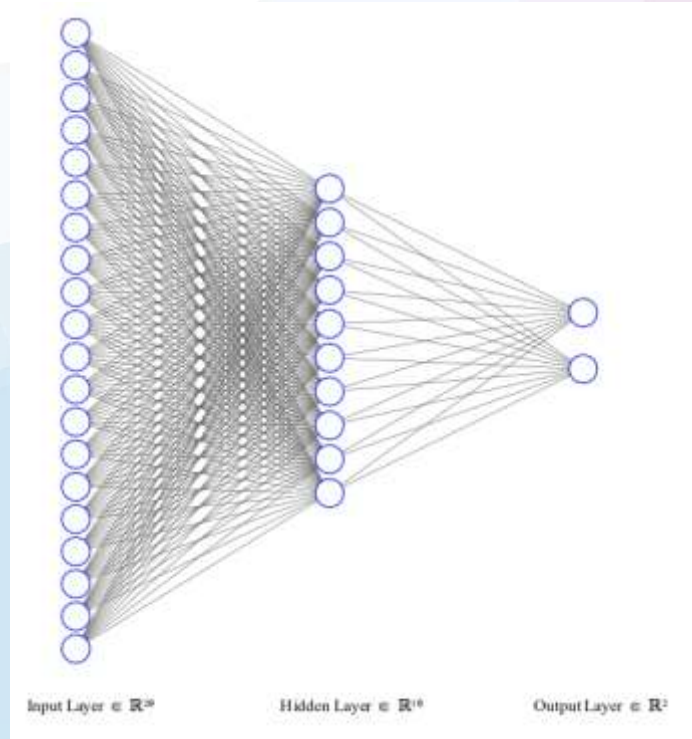
- Where is the „center of mass” such network?
- First part defines the most primitives schemas, but it is large and difficult to learn.

Important observations

- It is always easier to extract information from previously annotated data.
- We can reuse some derivative features for other tasks
 - Will it make big difference if we distinguish cats from dogs or big from small animals?

How can we create such model?

- Let's data speak for itself
 - Self supervised learning
 - Unsupervised learning



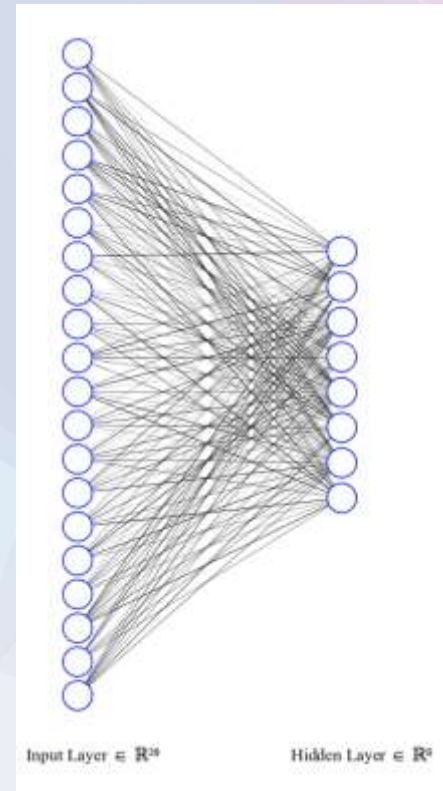
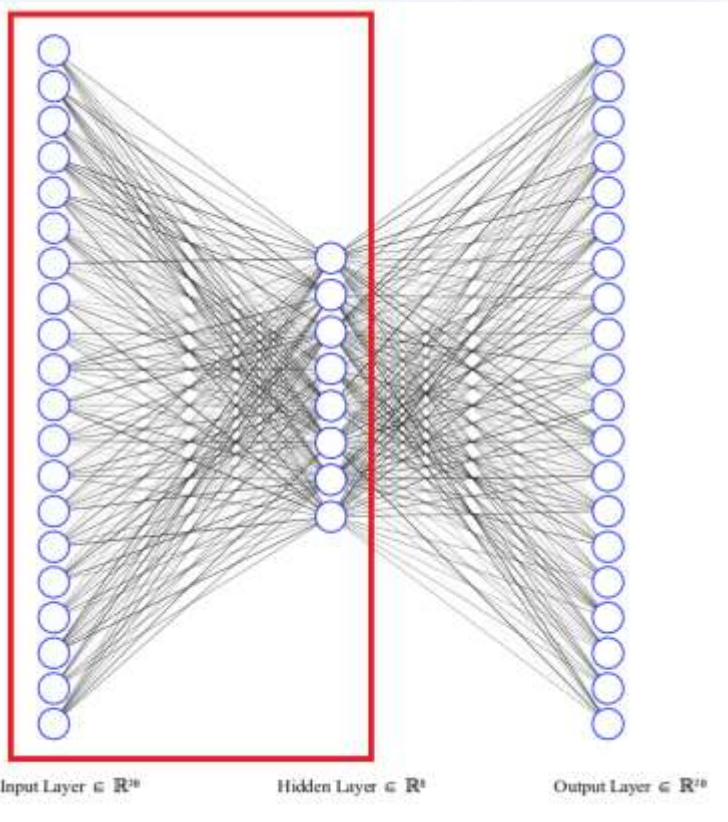
How can we create such model?

- We have model, what now?

How transfer learning works

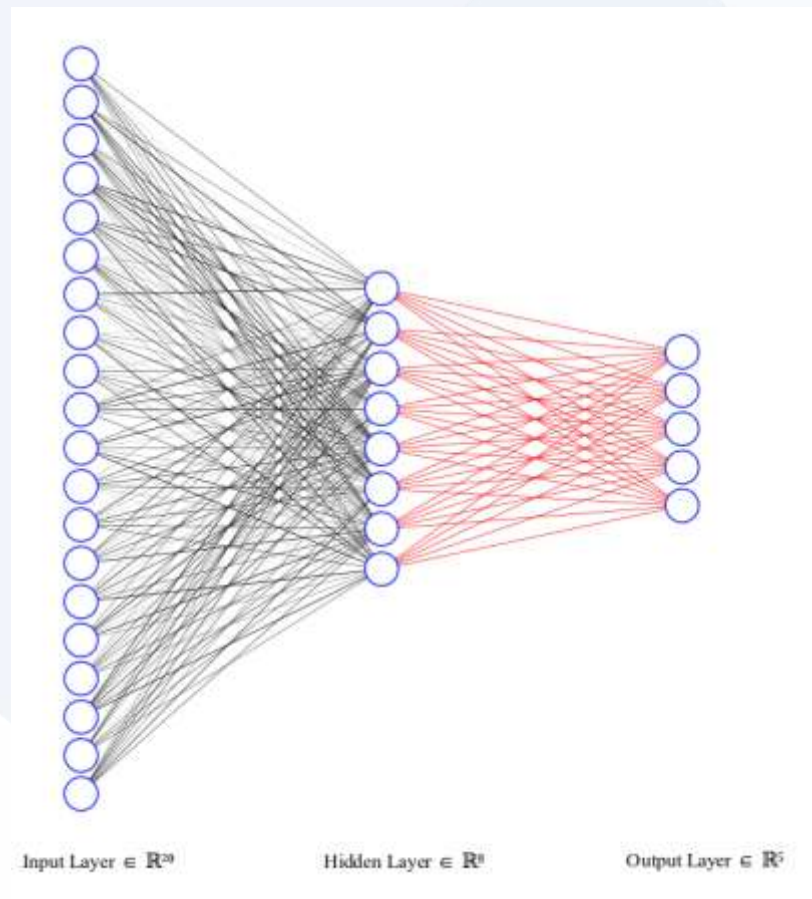
We pick a part of network.

That is our foundation part/model.

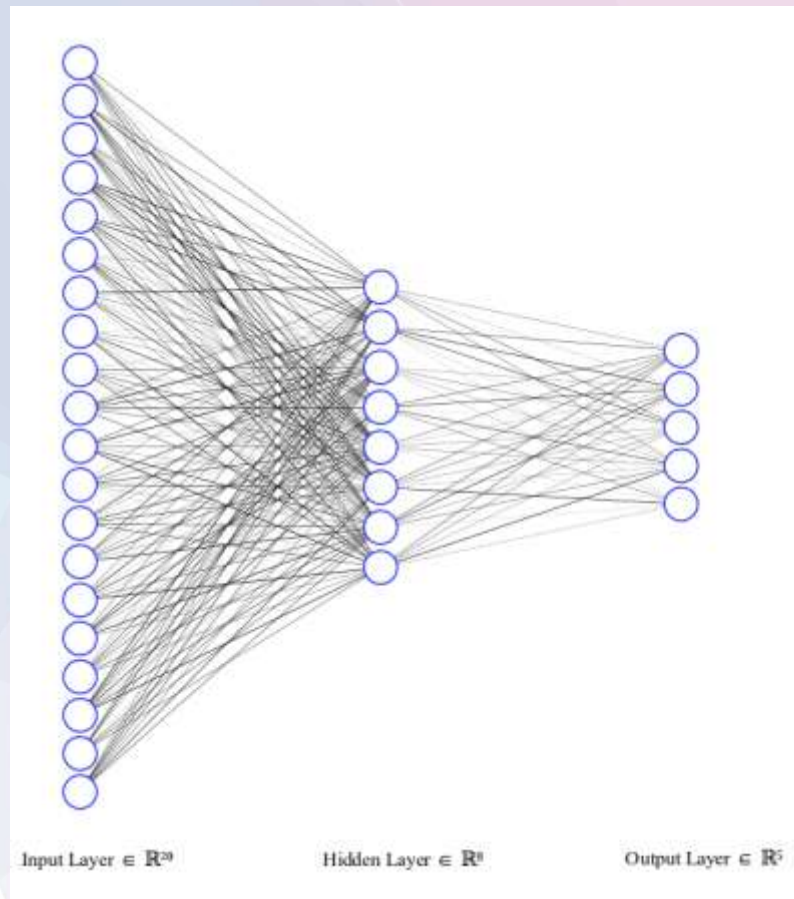


How fine tuning works

New model with attached next layer.



Tuned model with proper weights.



How they can help us?

- Reduce size of data required to training
- Reduced time required for training
- We can share our work without sharing data
- We can create robust base for other experiments
- Finally one task, by synergy, can improve other tasks.
- A few foundation models could solve many different tasks in whole organization.

Why they appear in context of language models?

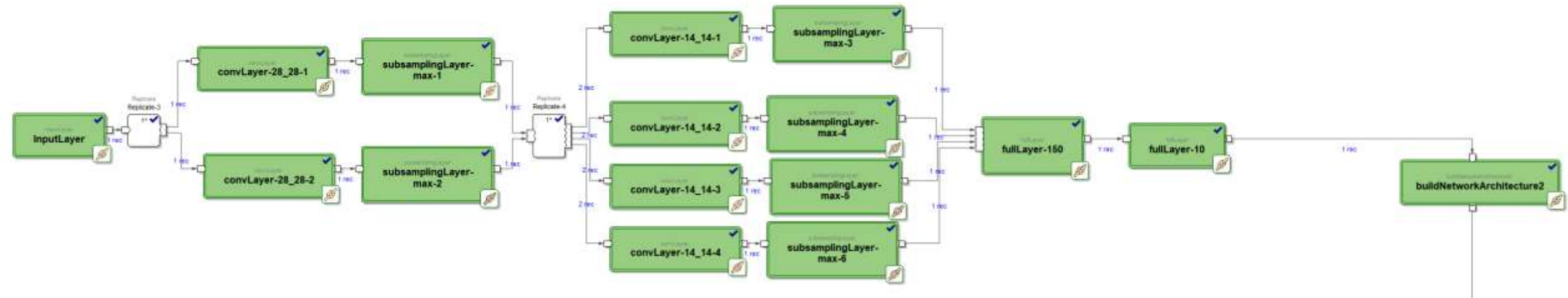
- Large amounts of data (really large, not just list of transactions in bank)
- Not possible to share corpora

Simple graphical tool example with mnist digits data set

How we can handle such tasks in a more convenient way?

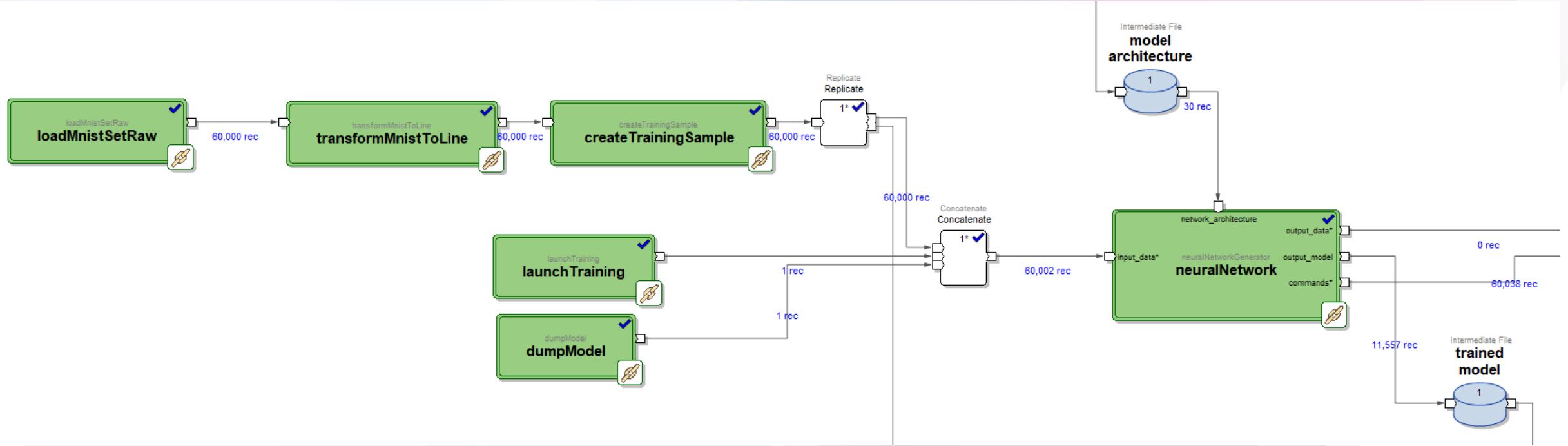
Simple graphical tool example with mnist digits data set

We create architecture network



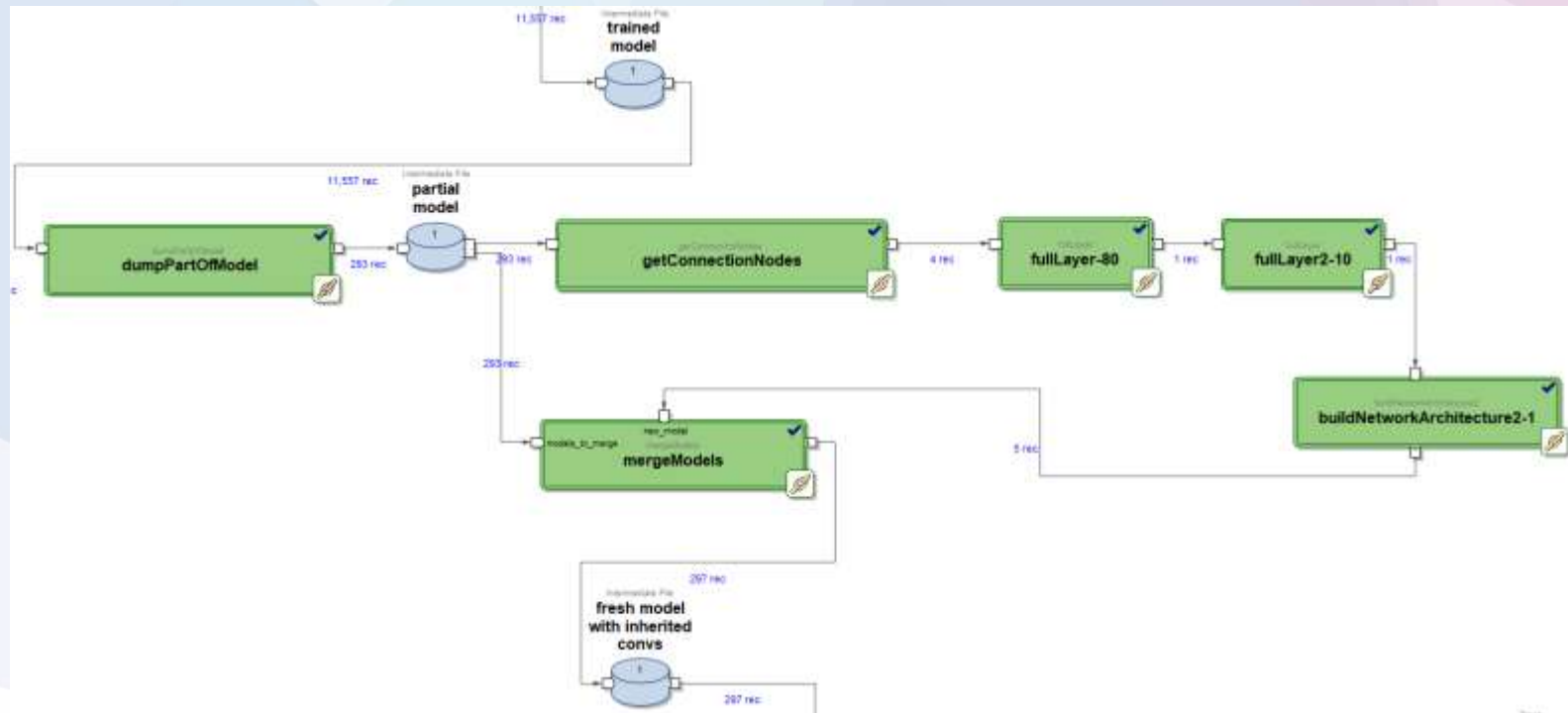
Simple graphical tool example with mnist digits data set

We prepare data and launch training



Simple graphical tool example with mnist digits data set

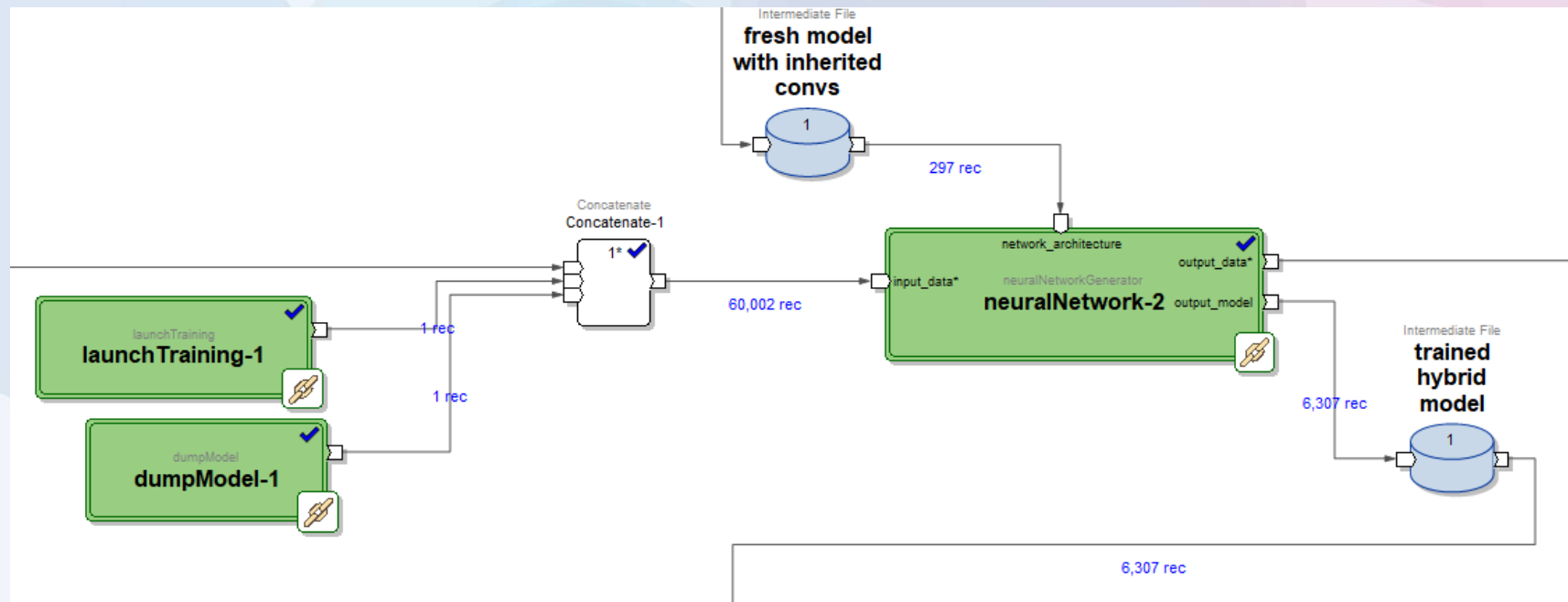
- We pick trained model and take only convolutive parts of network.
- We create new architecture of network.



Simple graphical tool example with mnist digits data set

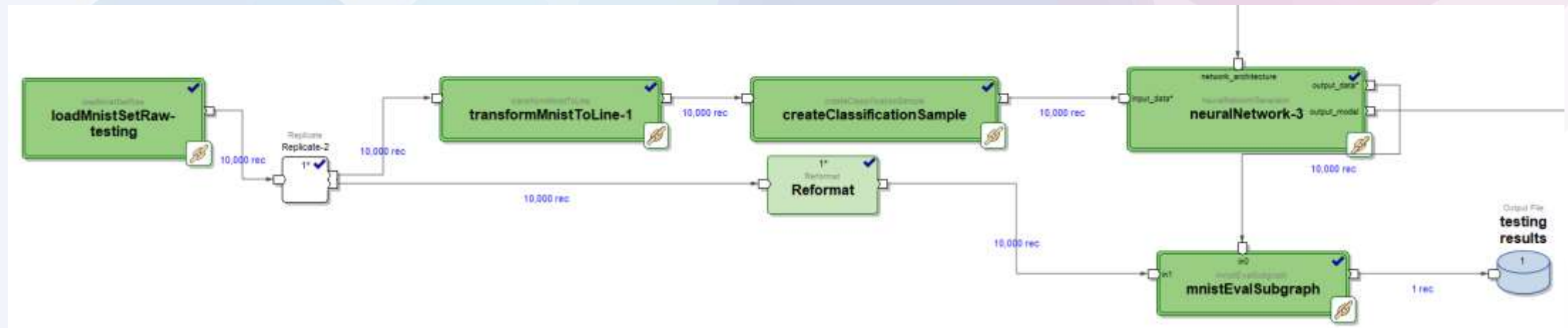
We train newly created network and launch only one training epoch (and only for newly attached layers).

Due to the pretrained first layers, training is much more efficient and many times faster.



Simple graphical tool example with mnist digits data set

Finally we evaluate our new fine-tuned model



View Data: testing results

	total	hit
1	10000	9658
[EOF]		

Tip of the iceberg

In terms of modifications, variants and improvements we are only restricted by our imagination. So make your own experiments.

Ab INITIO

Thank you

