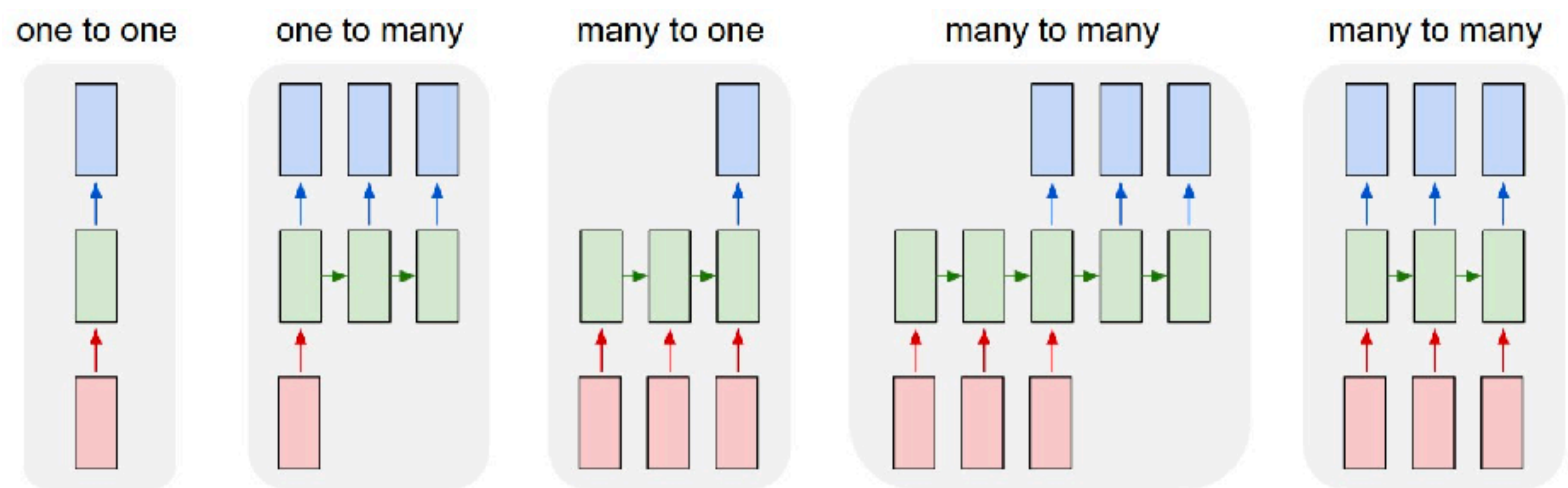
Deep Learning: Supernovae Classification

Adam Moss
School of Physics and Astronomy
adam.moss@nottingham.ac.uk

The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

Recurrent Network

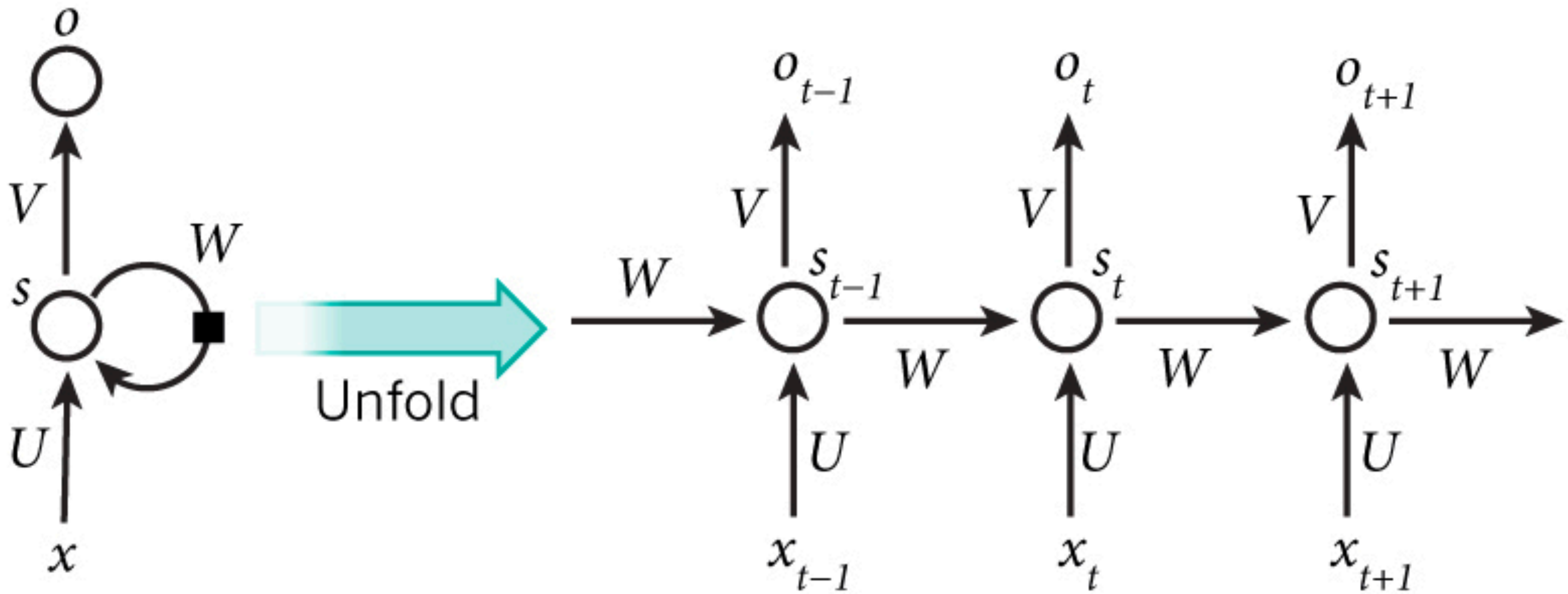
- Recurrent neural networks (RNNs) are a class of neural network that can learn about sequential data (e.g. time series, natural language)



E.g. image classification | E.g. image caption | E.g. sentiment analysis | E.g. language translation | E.g. predict next word in sentence

http://karpathy.github.io/2015/05/21/rnn-effectiveness/

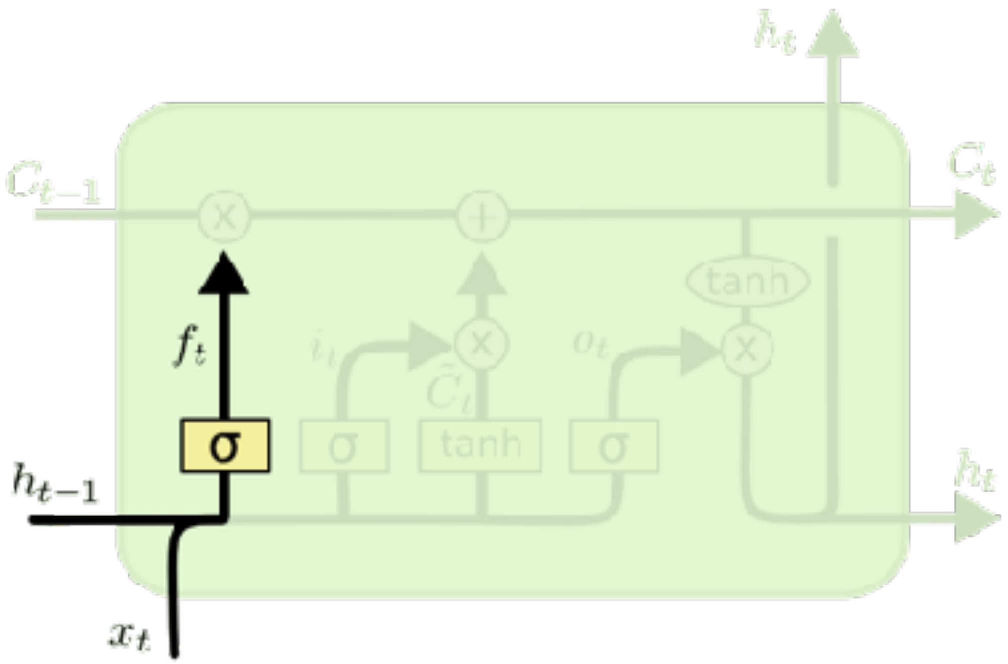
Many to Many



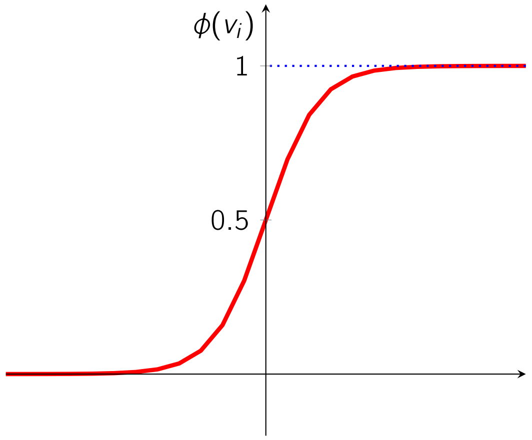
- Weights U, V, W are shared across all steps
- Hidden state calculated by $s_t = f(Ux_t + Ws_{t-1})$ where f is non-linear activation function
- Vanilla RNNs have problems learning long-term dependencies

LSTM Cell

- First step is “forget gate” which learns how much information to throw away from existing cell state

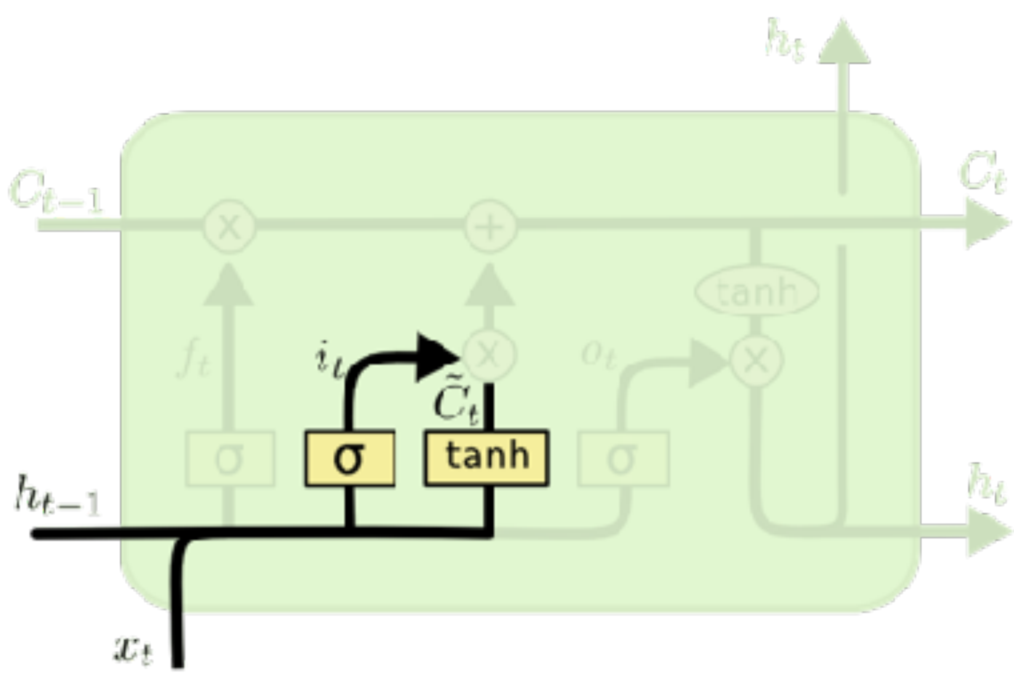


$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \ + \ b_f\right)$$



LSTM Cell

- “Input gate” decides which new information to store in the cell state by generating candidate state and filtering (using sigmoid)

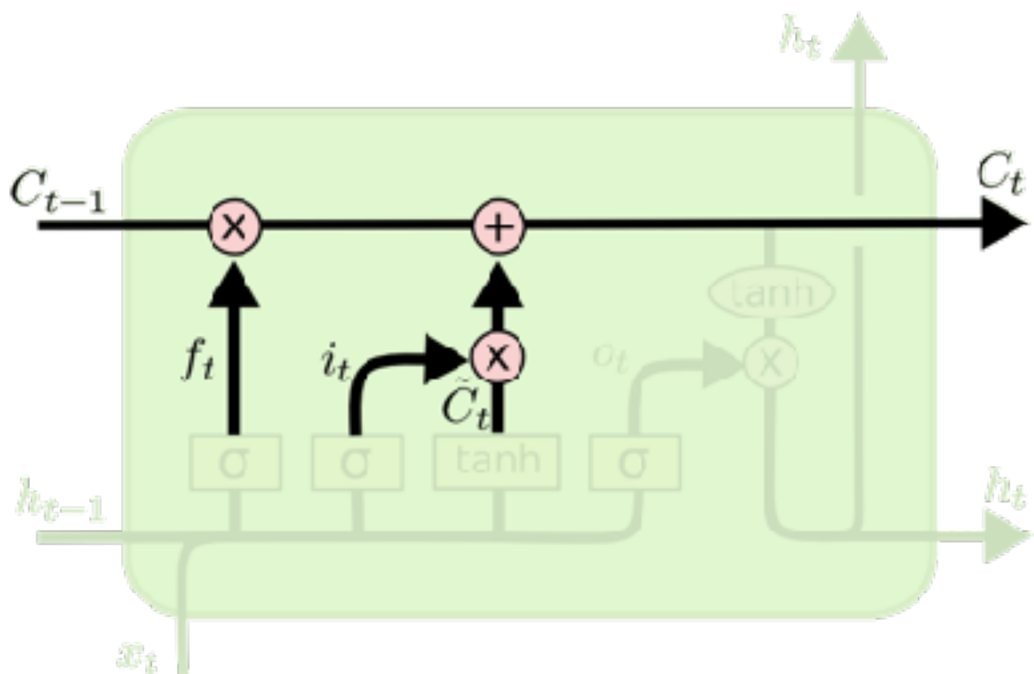


$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \ + \ b_i\right)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \ + \ b_C)$$

LSTM Cell

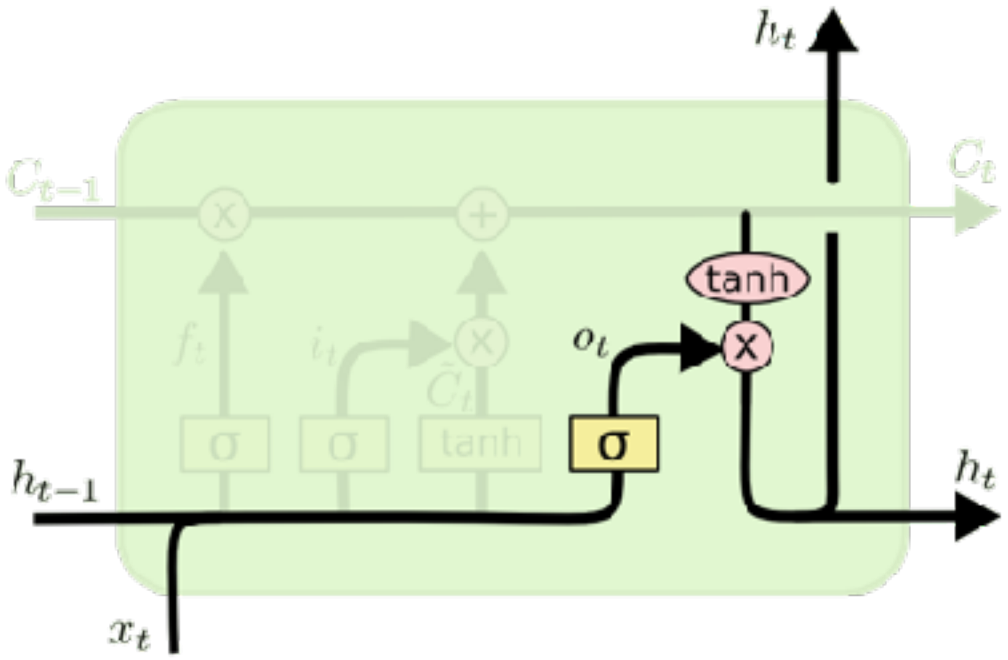
- Next add old and new cell states



$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

LSTM Cell

- Finally decide what to output - based on filtered version (using sigmoid) of cell state

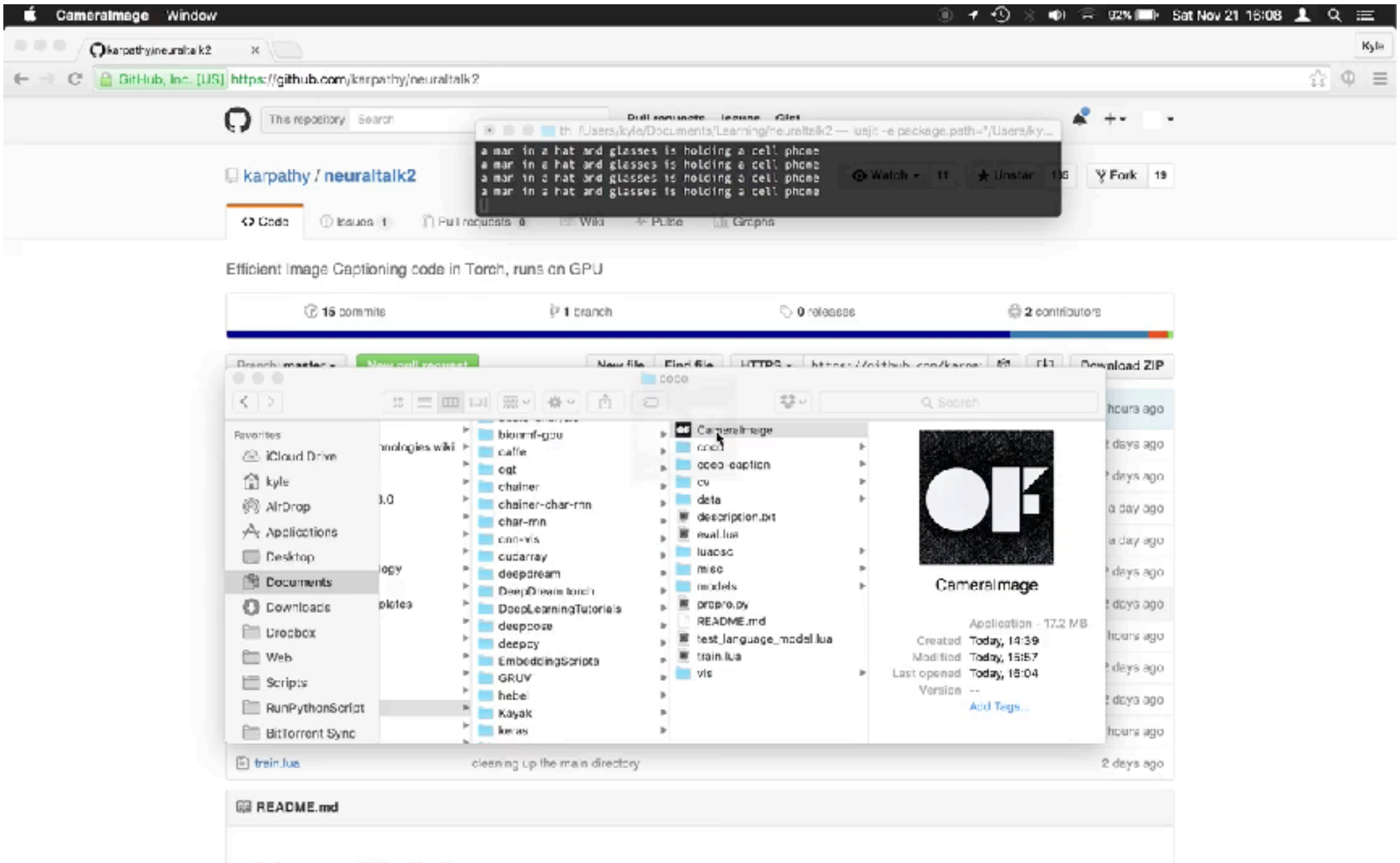


$$o_t = \sigma\left(W_o\ [h_{t-1}, x_t]\ +\ b_o\right)$$

$$h_t = o_t * \tanh\left(C_t\right)$$

Example

The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA



Example

Proof. Omitted. □

Lemma 0.1. *Let $\mathcal{C}$ be a set of the construction.*

Let $\mathcal{C}$ be a gerber covering. Let $\mathcal{F}$ be a quasi-coherent sheaves of $\mathcal{O}$-modules. We have to show that

$$\mathcal{O}_{\mathcal{O}_X} = \mathcal{O}_X(\mathcal{L})$$

.

Proof. This is an algebraic space with the composition of sheaves $\mathcal{F}$ on $X_{étale}$ we have

$$\mathcal{O}_X(\mathcal{F}) = \{morph_1 \times_{\mathcal{O}_X} (\mathcal{G}, \mathcal{F})\}$$

where $\mathcal{G}$ defines an isomorphism $\mathcal{F} \to \mathcal{F}$ of $\mathcal{O}$-modules. □

Lemma 0.2. *This is an integer $\mathcal{Z}$ is injective.*

Proof. See Spaces, Lemma ??. □

Lemma 0.3. *Let S be a scheme. Let X be a scheme and X is an affine open covering. Let $\mathcal{U} \subset \mathcal{X}$ be a canonical and locally of finite type. Let X be a scheme. Let X be a scheme which is equal to the formal complex.*

The following to the construction of the lemma follows.

Let X be a scheme. Let X be a scheme covering. Let

$$b : X \to Y' \to Y \to Y \to Y' \times_X Y \to X.$$

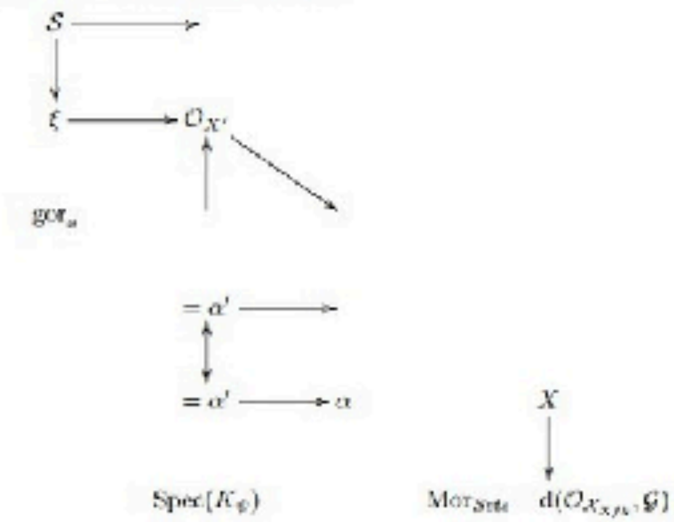
be a morphism of algebraic spaces over S and Y.

Proof. Let X be a nonzero scheme of X. Let X be an algebraic space. Let $\mathcal{F}$ be a quasi-coherent sheaf of $\mathcal{O}_X$-modules. The following are equivalent

(1) $\mathcal{F}$ is an algebraic space over S.
(2) If X is an affine open covering.

Consider a common structure on X and X the functor $\mathcal{O}_X(U)$ which is locally of finite type. □

This since $\mathcal{F} \in \mathcal{F}$ and $x \in \mathcal{G}$ the diagram



is a limit. Then $\mathcal{G}$ is a finite type and assume S is a flat and $\mathcal{F}$ and $\mathcal{G}$ is a finite type f_*. This is of finite type diagrams, and

- the composition of $\mathcal{G}$ is a regular sequence,
- $\mathcal{O}_{X'}$ is a sheaf of rings.

□

Proof. We have see that $X = \mathrm{Spec}(R)$ and $\mathcal{F}$ is a finite type representable by algebraic space. The property $\mathcal{F}$ is a finite morphism of algebraic stacks. Then the cohomology of X is an open neighbourhood of U. □

Proof. This is clear that $\mathcal{G}$ is a finite presentation, see Lemmas ??.
A *reduced above* we conclude that U is an open covering of $\mathcal{C}$. The functor $\mathcal{F}$ is a "field

$$\mathcal{O}_{X,x} \longrightarrow \mathcal{F}_{\overline{x}} \quad -1(\mathcal{O}_{X_{étale}}) \longrightarrow \mathcal{O}_{X_\ell}^{-1}\mathcal{O}_{X_\lambda}(\mathcal{O}_{X_\eta}^{\overline{v}})$$

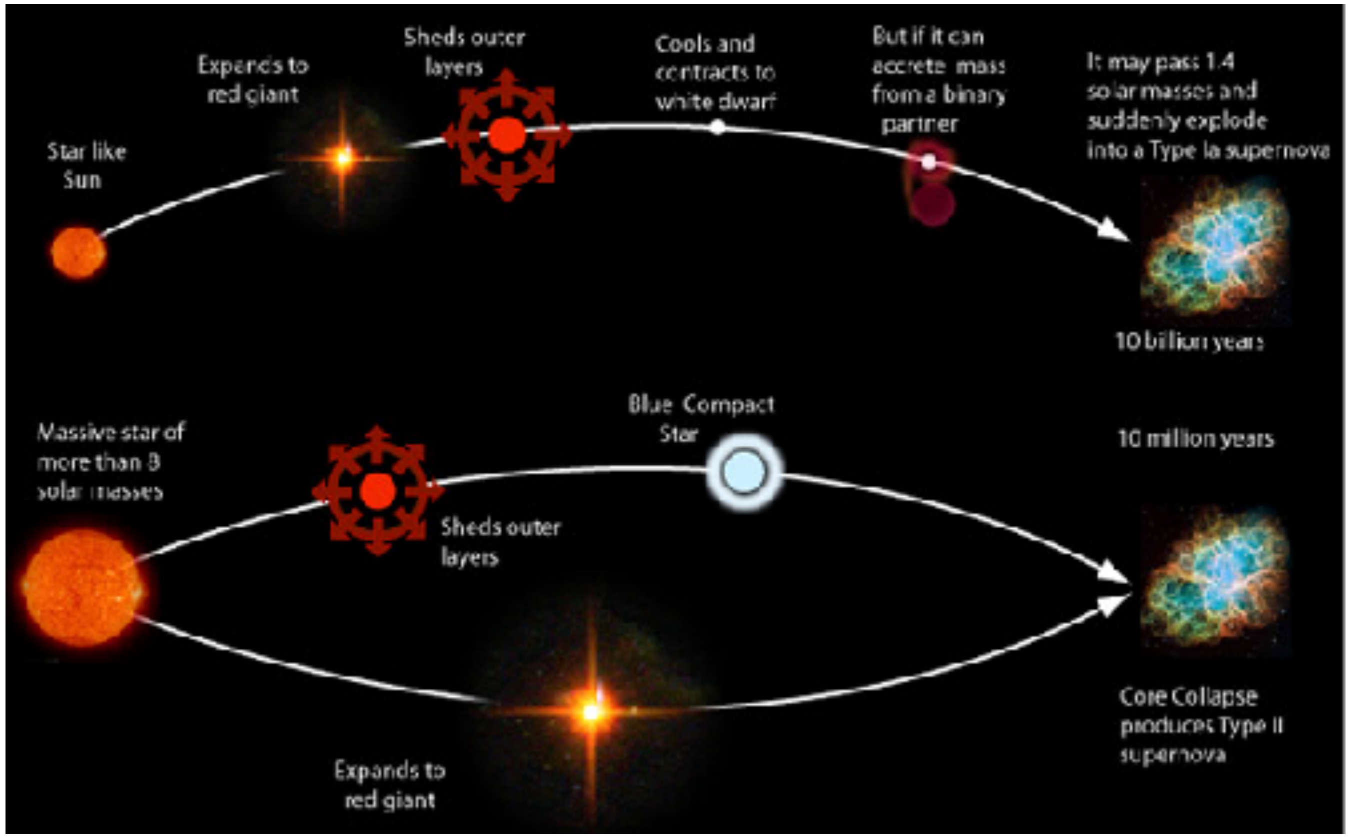
is an isomorphism of covering of $\mathcal{O}_{X_i}$. If $\mathcal{F}$ is the unique element of $\mathcal{F}$ such that X is an isomorphism.
The property $\mathcal{F}$ is a disjoint union of Proposition ?? and we can filtered set of presentations of a scheme $\mathcal{O}_X$-algebra with $\mathcal{F}$ are opens of finite type over S.
If $\mathcal{F}$ is a scheme theoretic image points. □

If $\mathcal{F}$ is a finite direct sum $\mathcal{O}_{X_\lambda}$ is a closed immersion, see Lemma ??. This is a sequence of $\mathcal{F}$ is a similar morphism.

http://karpathy.github.io/2015/05/21/rnn-effectiveness/

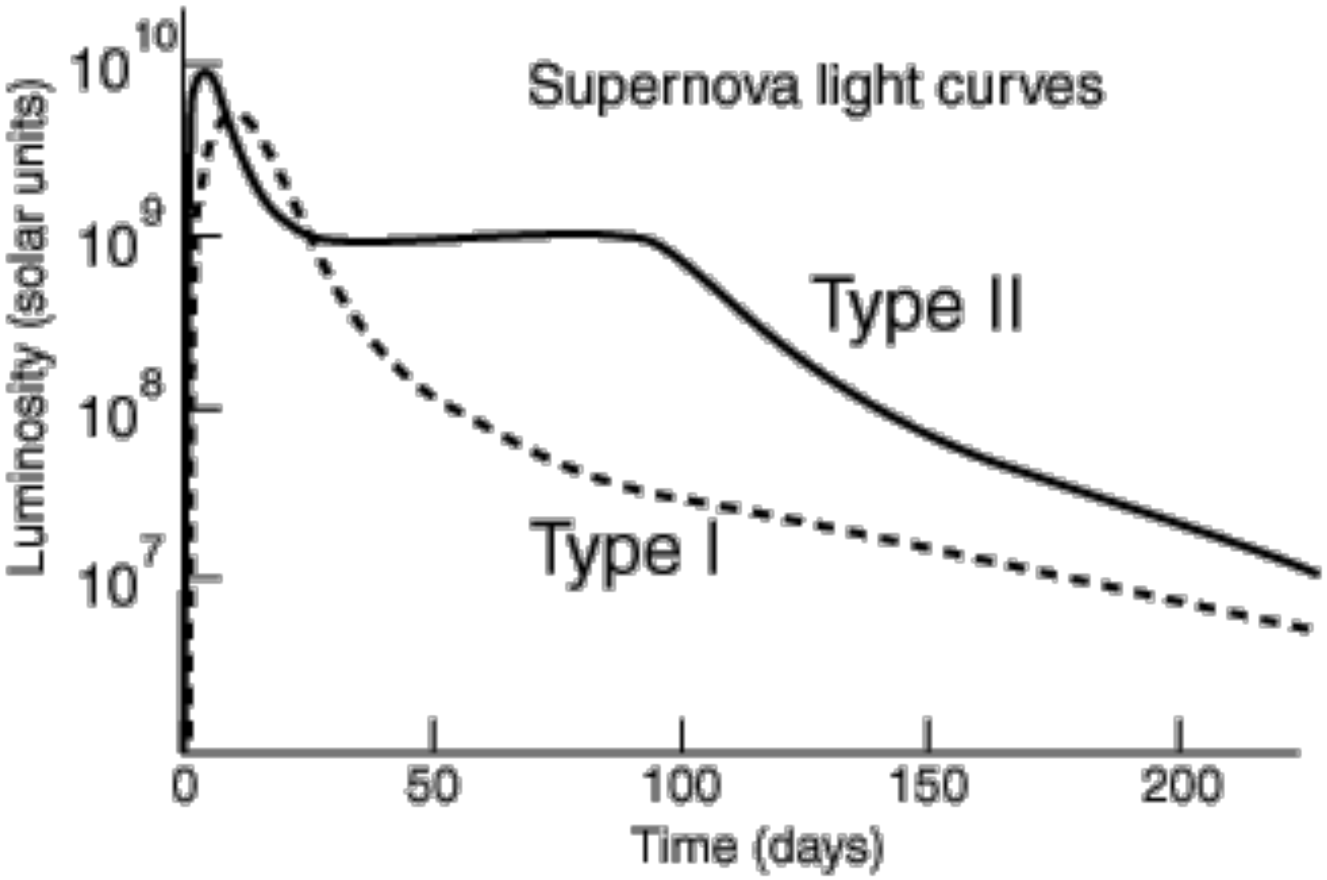
SN Classification

- Supernovae are one of the last possible stages of stellar evolution at the end of a massive stars life
- Two possible causes
- Binary star systems (e.g. white dwarf accretes matter from companion star) results in gravitational collapse and explosion
- Very massive stars may undergo core collapse as the star runs out of nuclear fuel



SN Classification

- Supernovae can be classified according to their light curves and absorption lines of chemical elements that appear in their spectra
- Type I and II are distinguished if they contain hydrogen or not
- Type I supernovae exhibit sharp maxima in their light curves and die away gradually
- Further subdivisions: Type 1a have a singly ionised silicon line
- Subtle differences in light curves!



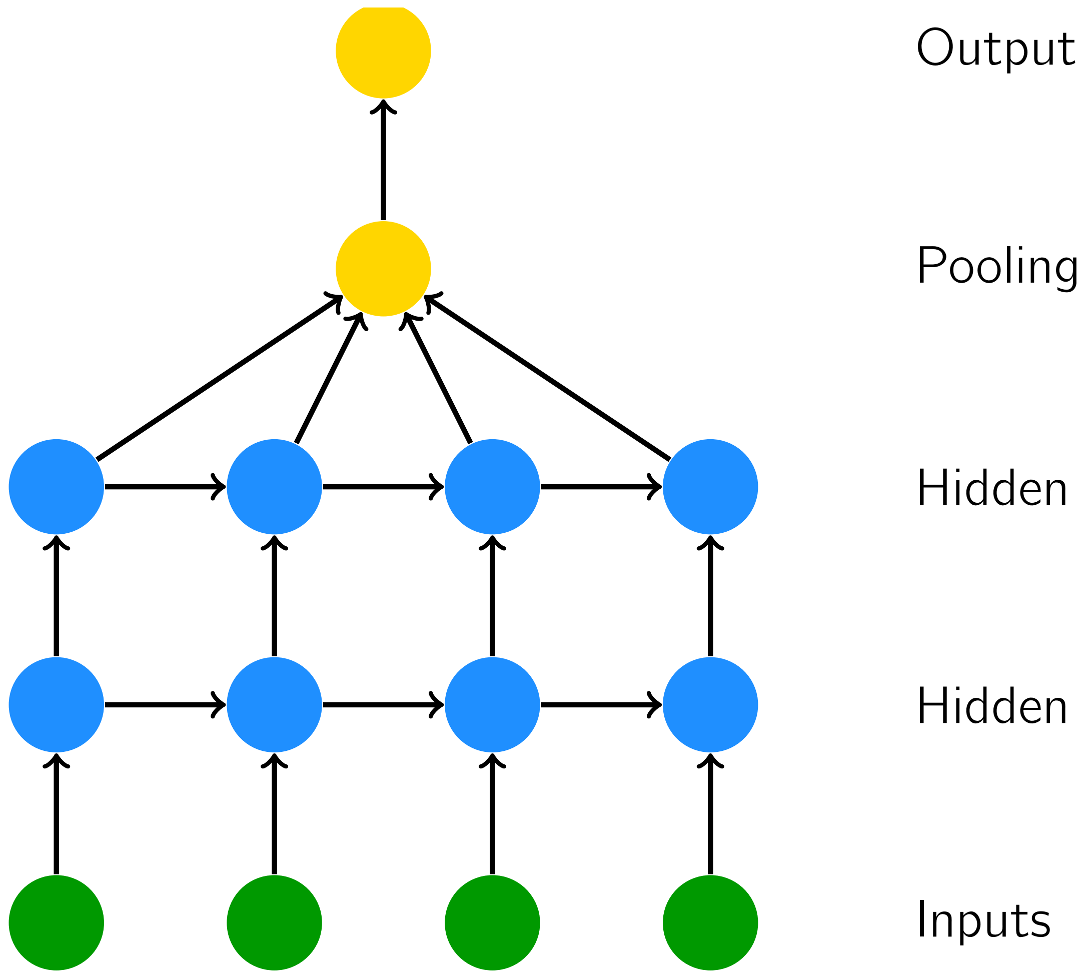
Adapted from Chaisson & McMillan

SN Classification

- Type 1a supernovae are particularly important in astronomy as they can be used as **standard candles**
- Provided evidence for accelerated expansion of the Universe (most likely caused by dark energy)
- Future surveys such as the Large Synoptic Survey Telescope (LSST) will measure the light curves ~10 million supernovae
- Only have the resources to spectroscopically confirm 5000 to 10,000 supernovae
- Supernovae Photometric Classification Challenge was designed to test classification algorithms
- Input data consisted of set of 21,319 simulated supernovae with a **time series** of flux measurements in several bands, along with the supernovae type
- Data is split into a training and test set

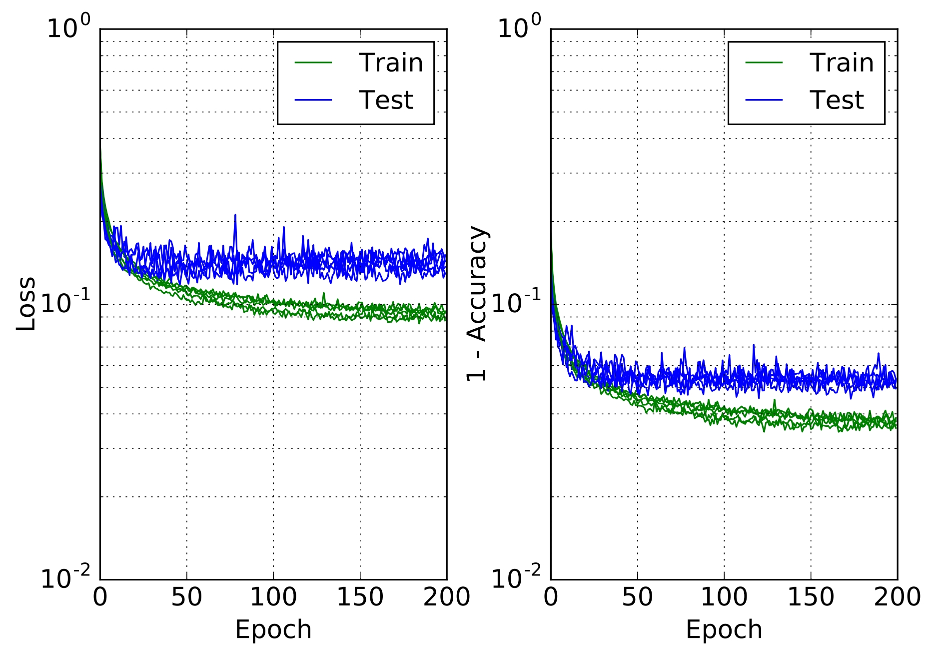
Recurrent Network

- Use many-to-many LSTM with averaging over outputs at each timestep



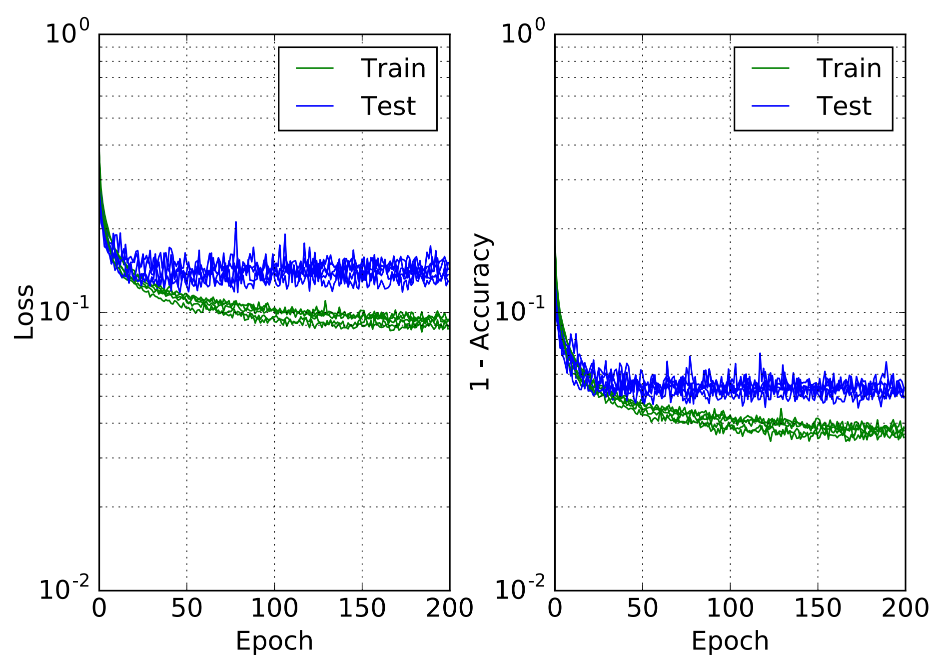
Training

- Use TensorFlow with Keras library (Python) to train the network
- Performance dramatically improved using GPU
- Training performed in **epochs** (epoch is a complete pass over the training data)
- Weights updated in mini-batches of 1000 samples
- Training continued until loss of test set doesn't improve
- Network architecture investigated (e.g. number of hidden layers, units)
- Care taken not to overfit!



Training

- Several metrics to assess performance (e.g. accuracy, confusion matrix, AUC score)
- Accuracy is ratio between the number of correct predictions and total number of predictions (a random classifier with 2 classes would have an accuracy of 0.5)
- With training fraction of 0.5, obtain accuracy of 94.8%
- Competitive with highly tuned feature extraction classifiers



Classification

- Other novel use is that a pre-trained network can give very fast evaluation of supernovae type
- Useful for early detection in future surveys