ELEC 677: Introduction to Deep Machine Learning

Ankit B. Patel
Baylor College of Medicine (Neuroscience Dept.)
Rice University (ECE Dept.)
8-23-2016

About Me

- **Education**
  - Ph.D in Applied Mathematics/Computer Science (Harvard 2008)
- **Industry (building real-time inference systems)**
  - MIT Lincoln Laboratory: Ballistic Missile Defense (2 years)
  - High-Frequency Trading (4 years): fast inference on extremely large datasets
- **Return to Academia**
  - Postdoc (Rich Baraniuk): theory of deep learning
  - New Faculty at Baylor College of Medicine (Neuroscience), joint with Rice ECE: check out ankitlab.co for more details about my lab's mission at the intersection of deep learning and computational neuroscience

Deep Learning:
A Short Preview

Why do we need Deep Learning?

[Girshick et al., CVPR 2014]

Key Challenge: Object recognition (and sensory perception in general) is plagued by large amounts of **nuisance variation**.

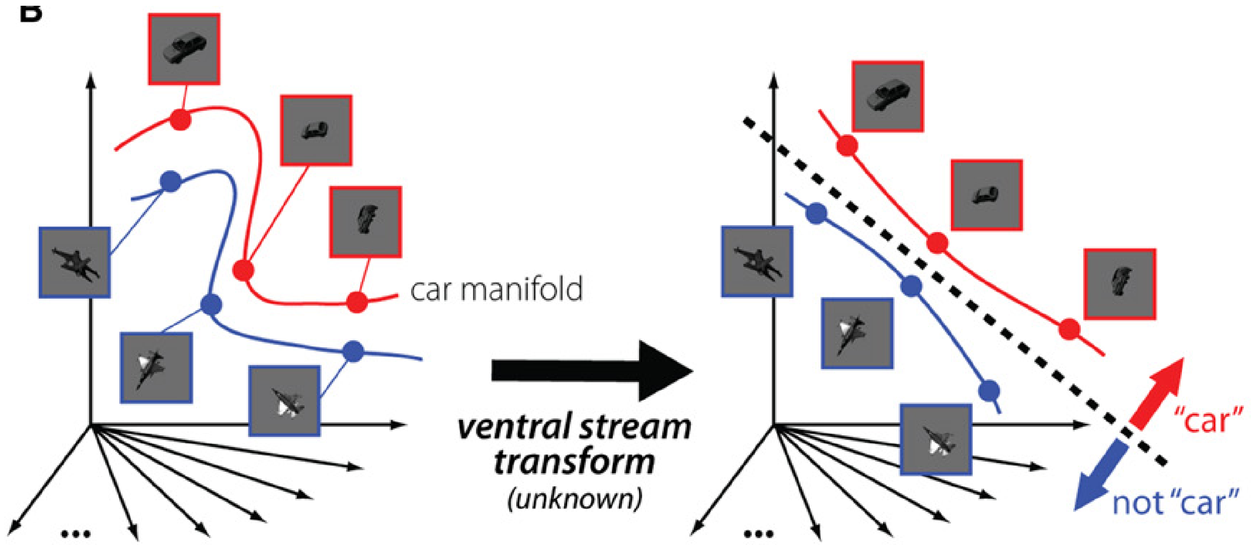
- **Nuisance Variation**: affects sensory input (image) but not the task target (object class)
- Ex: Object Recognition, Nuisances = changes in location, pose, viewpoint, lighting, expression, ...
- Ex: Speech Recognition, Nuisances = changes in pitch, volume, pace, accent, ...
- Nuisance variables are task-dependent and can be implicit

Why do we need Deep Learning?

Problem: *How to deal with nuisance variation in the input?*

Solution: Build representations that are

- **Selective**: Sensitive to task-relevant (target) features
- **Invariant:** Robust to task-irrelevant (nuisance) features
- **Multi-task:** Useful for many different tasks



DiCarlo, J. J. et al. How does the brain solve visual object recognition? Neuron (2012).

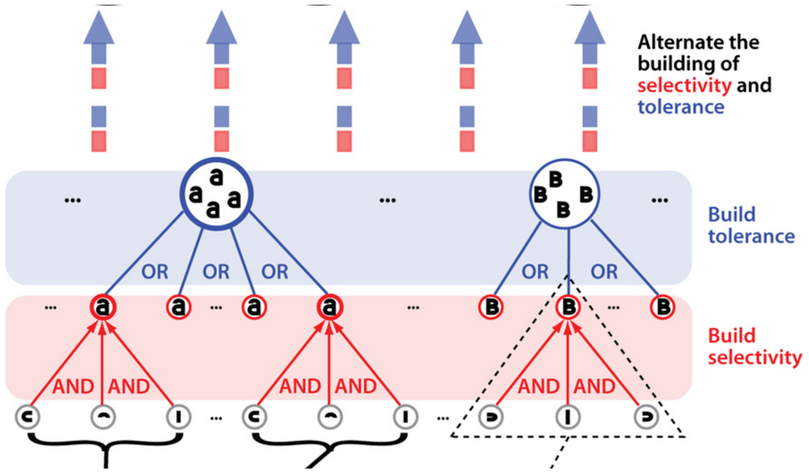
The Holy Grail of Machine Learning

Learn a **disentangled representation**:
one that factors out variation in the sensory input
into meaningful intrinsic degrees of freedom.

Why do we need Deep Learning?

Potential Solution: Look to the Brain for guidance.

- Hubel and Wiesel's discovery of simple/complex cells and their special properties of *selectivity* and *tolerance/invariance*



DiCarlo, J. J. et al. How does the brain solve visual object recognition? Neuron (2012).

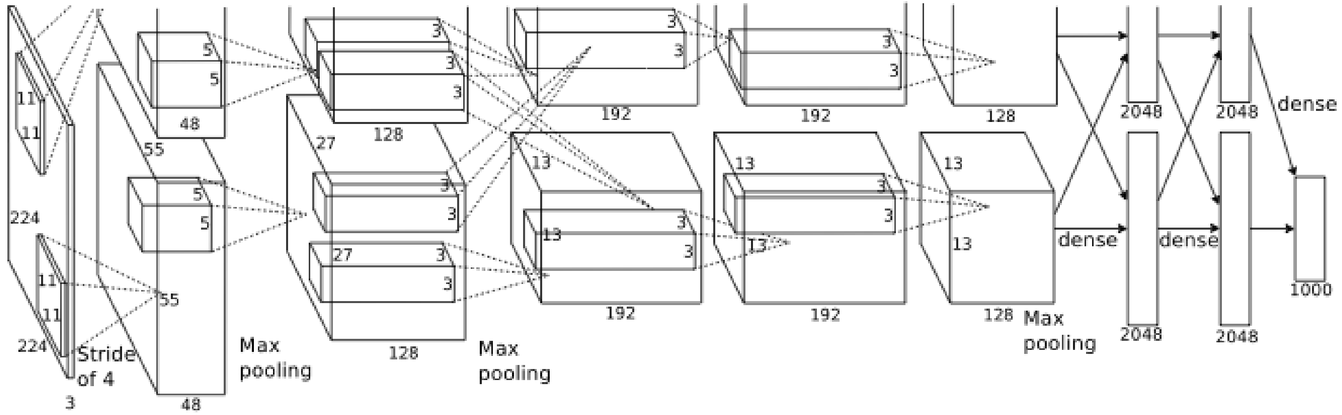
Key Inspiration from Neuroscience

Build up feature selectivity and tolerance over multiple layers in a hierarchy $\Rightarrow$
ML architectures: Neocognitron, HMAX, SIFT, and modern **Deep Convnets**

Neural Networks

- Takes in inputs and returns outputs
- Layers of processing: alternates between linear and nonlinear transformations typically
- High expressive power / can be trained to learn complex functions
- Inspired by the brain

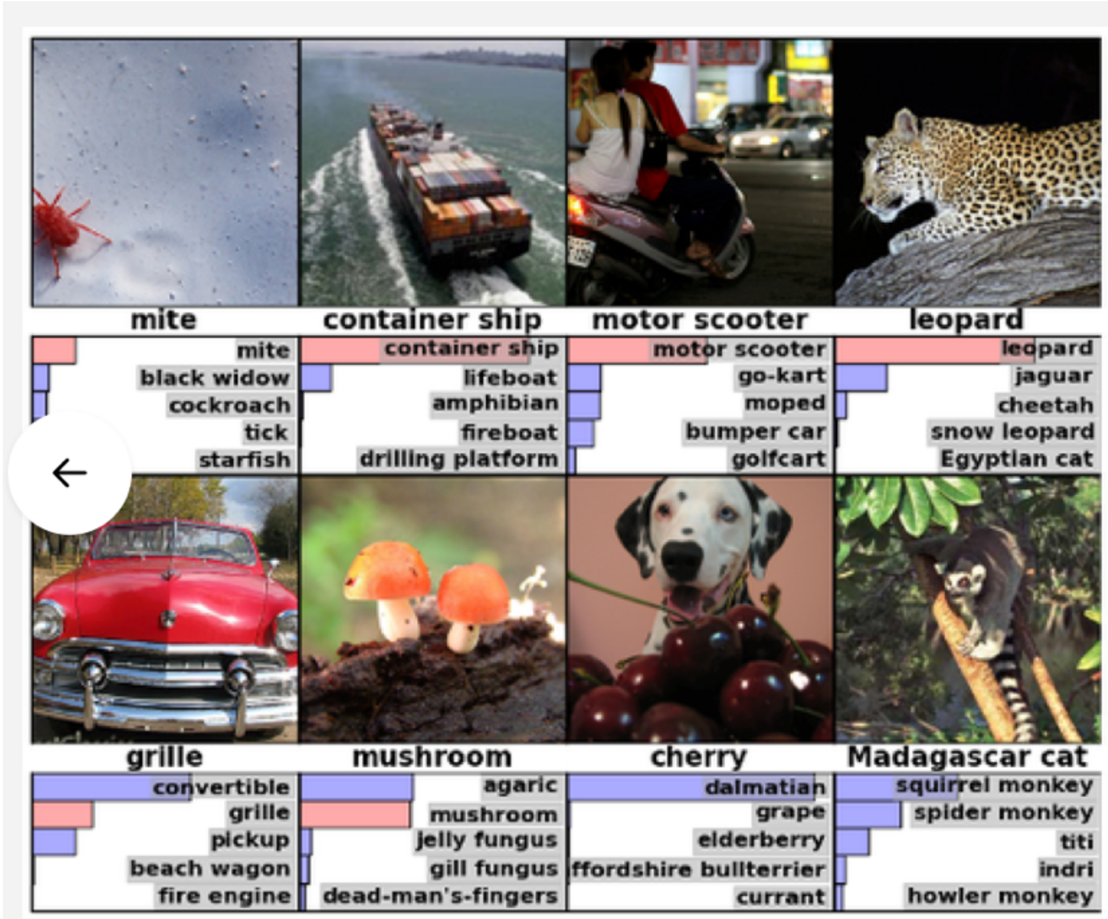
Object Recognition with Convnets



A. Krizhevsky et al. ImageNet classification with deep convolutional neural networks (NIPS 2012)

- **Deep Convnets**
  - 2012: Krizhevsky et al advanced state-of-the-art in object recognition in the ImageNet Challenge (1.2 million labeled images of objects)
  - Subsequently benchmarks in **many other vision tasks** were pushed forward many years $\Rightarrow$ **Transfer Learning**
  - Recently, Google's and MSR's latest DCNs have achieved **95% accuracy**, with **superhuman** performance in most categories
  - Deployed commercially in Google and Baidu Personal Image Search

Object Recognition with Convnets



Facial Recognition/Verification

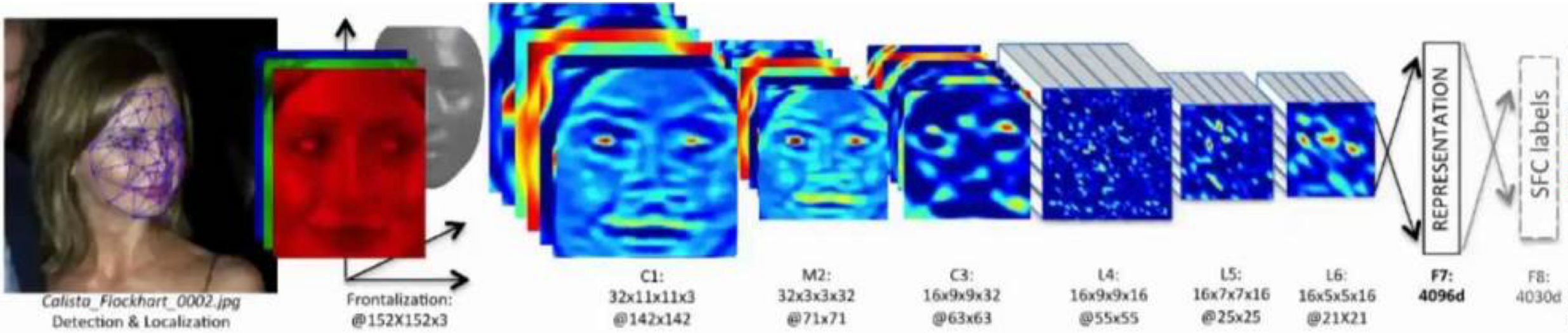
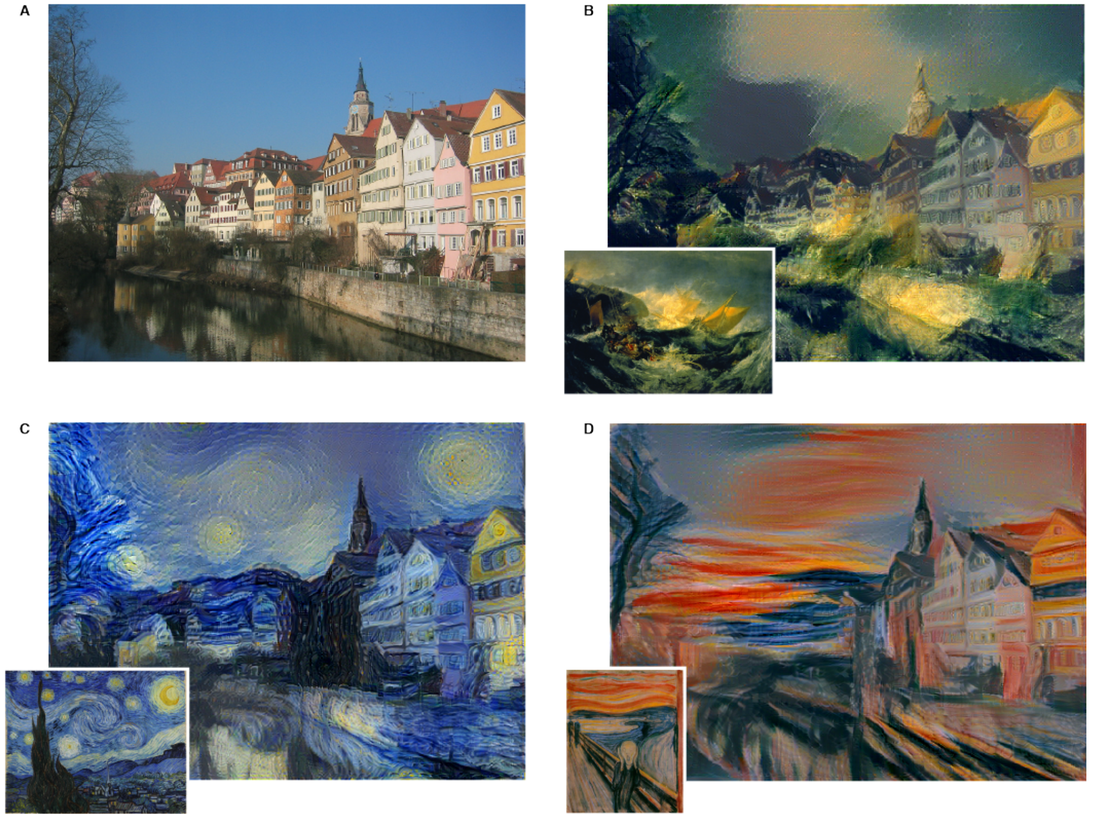


Figure 2. **Outline of the *DeepFace* architecture**. A front-end of a single convolution-pooling-convolution filtering on the rectified input, followed by three locally-connected layers and two fully-connected layers. Colors illustrate feature maps produced at each layer. The net includes more than 120 million parameters, where more than 95% come from the local and fully connected layers.

Face Recognition: Detect → Alignment → Represent → Classify

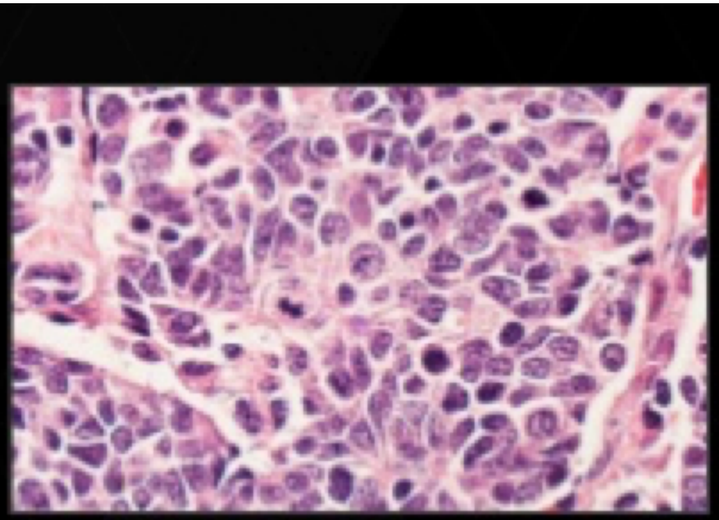
Deep Art: Combining Content and Style from Different Images

- Coarse-scale **Content** from one image, Fine-scale **Style** from another image
- **Observation:** DCNs learn sophisticated multi-scale representations
- **Theoretical Result:** Mathematical formulation of separation of length scales $\Rightarrow$ levels of abstraction



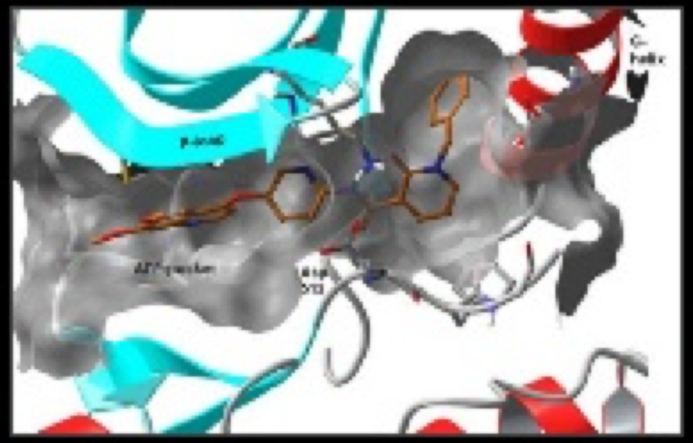
L. Gatys, A. Ecker, M. Bethge
A Neural Algorithm of Artistic Style (ArXiV 2015: eprint arXiv:1508.06576)

Many Medical Applications



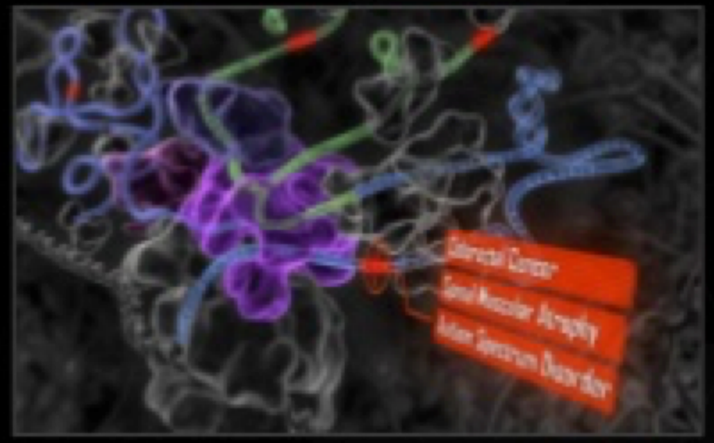
Detecting Mitosis in Breast Cancer Cells

— IDSIA



Predicting the Toxicity of New Drugs

— Johannes Kepler University



Understanding Gene Mutation to Prevent Disease

— University of Toronto

Playing Video Games

https://www.youtube.com/watch?v=V1eYniJ0Rnk

Playing AlphaGo

Self-Driving Cars

https://www.youtube.com/watch?v=QpWTyFIUvYk

Deep Sensorimotor Learning for Robotics

https://www.youtube.com/watch?v=Es83Co_Vz78

Generative Models for Natural Images

(a) Varying c_1 on InfoGAN (Digit type)

(c) Varying c_2 from -2 to 2 on InfoGAN (Rotation)

Generating Shakespeare

```
PANDARUS:
Alas, I think he shall be come approached and the day
When little srain would be attain'd into being never fed,
And who is but a chain and subjects of his death,
I should not sleep.

Second Senator:
They are away this miseries, produced upon my soul,
Breaking and strongly should be buried, when I perish
The earth and thoughts of many states.

DUKE VINCENTIO:
Well, your wit is in the care of side and that.

Second Lord:
They would be ruled after this chamber, and
my fair nues begun out of the fact, to be conveyed,
Whose noble souls I'll have the heart of the wars.

Clown:
Come, sir, I will make did behold your worship.

VIOLA:
I'll drink it.
```

Generating Wiki Markup

```
{ { cite journal | id=Cerling Nonforest Department|format=Newlymeslated|none } }
''www.e-complete''.

'''See also''': [[List of ethical consent processing]]

== See also ==
*[[Iender dome of the ED]]
*[[Anti-autism]]

===[[Religion|Religion]]===
*[[French Writings]]
*[[Maria]]
*[[Revelation]]
*[[Mount Agamul]]

== External links==
* [http://www.biblegateway.nih.gov/entrepre/ Website of the World Festival. The labour

==External links==
* [http://www.romanology.com/ Constitution of the Netherlands and Hispanic Competition
```

Generating Linux Source Code

```
static void do_command(struct seq_file *m, void *v)
{
  int column = 32 << (cmd[2] & 0x80);
  if (state)
    cmd = (int)(int_state ^ (in_8(&ch->ch_flags) & Cmd) ? 2 : 1);
  else
    seq = 1;
  for (i = 0; i < 16; i++) {
    if (k & (1 << 1))
      pipe = (in_use & UMXTHREAD_UNCCA) +
        ((count & 0x00000000ffffffff8) & 0x000000f) << 8;
    if (count == 0)
      sub(pid, ppc_md.kexec_handle, 0x20000000);
    pipe_set_bytes(i, 0);
  }
  /* Free our user pages pointer to place camera if all dash */
  subsystem_info = &of_changes[PAGE_SIZE];
  rek_controls(offset, idx, &soffset);
  /* Now we want to deliberately put it to device */
  control_check_polarity(&context, val, 0);
  for (i = 0; i < COUNTER; i++)
    seq_puts(s, "policy ");
}
```

Generating Algebraic Topology

For $\bigoplus_{n=1,\ldots,m}$ where $\mathcal{L}_{m_\bullet} = 0$, hence we can find a closed subset $\mathcal{H}$ in $\mathcal{H}$ and any sets $\mathcal{F}$ on X, U is a closed immersion of S, then $U \to T$ is a separated algebraic space.

Proof. Proof of (1). It also start we get

$$S = \mathrm{Spec}(R) = U \times_X U \times_X U$$

and the compariciy in the fibre product covering we have to prove the lemma generated by $\coprod Z \times_U U \to V$. Consider the maps M along the set of points Sch_{fppf} and $U \to U$ is the fibre category of S in U in Section, **??** and the fact that any U affine, see Morphisms, Lemma **??**. Hence we obtain a scheme S and any open subset $W \subset U$ in $Sh(G)$ such that $\mathrm{Spec}(R') \to S$ is smooth or an

$$U = \bigcup U_i \times_{S_i} U_i$$

which has a nonzero morphism we may assume that f_i is of finite presentation over S. We claim that $\mathcal{O}_{X,x}$ is a scheme where $x, x', s'' \in S'$ such that $\mathcal{O}_{X,x'} \to \mathcal{O}'_{X',x'}$ is separated. By Algebra, Lemma **??** we can define a map of complexes $\mathrm{GL}_{S'}(x'/S'')$ and we win. □

To prove study we see that $\mathcal{F}|_U$ is a covering of $\mathcal{X}'$, and $\mathcal{T}_i$ is an object of $\mathcal{F}_{X/S}$ for $i > 0$ and $\mathcal{F}_p$ exists and let $\mathcal{F}_i$ be a presheaf of $\mathcal{O}_X$-modules on $\mathcal{C}$ as a $\mathcal{F}$-module. In particular $\mathcal{F} = U/\mathcal{F}$ we have to show that

$$\widetilde{M}^\bullet = \mathcal{I}^\bullet \otimes_{\mathrm{Spec}(k)} \mathcal{O}_{S,s} - i_X^{-1}\mathcal{F})$$

is a unique morphism of algebraic stacks. Note that

$$\mathrm{Arrows} = (Sch/S)^{opp}_{fppf}, (Sch/S)_{fppf}$$

and

$$V = \Gamma(S, \mathcal{O}) \longmapsto (U, \mathrm{Spec}(A))$$

is an open subset of X. Thus U is affine. This is a continuous map of X is the inverse, the groupoid scheme S.

Proof. See discussion of sheaves of sets. □

The result for prove any open covering follows from the less of Example **??**. It may replace S by $X_{spaces,étale}$ which gives an open subspace of X and T equal to S_{Zar}, see Descent, Lemma **??**. Namely, by Lemma **??** we see that R is geometrically regular over S.

Lemma 0.1. *Assume (3) and (3) by the construction in the description.*

Suppose $X = \lim |X|$ *(by the formal open covering* X *and a single map* $\underline{Proj}_X(\mathcal{A}) = \mathrm{Spec}(B)$ *over* U *compatible with the complex*

$$Set(\mathcal{A}) = \Gamma(X, \mathcal{O}_{X,\mathcal{O}_X}).$$

When in this case of to show that $\mathcal{Q} \to \mathcal{C}_{Z/X}$ *is stable under the following result in the second conditions of (1), and (3). This finishes the proof. By Definition* **??** *(without element is when the closed subschemes are catenary. If* T *is surjective we may assume that* T *is connected with residue fields of* S*. Moreover there exists a closed subspace* $Z \subset X$ *of* X *where* U *in* X' *is proper (some defining as a closed subset of the uniqueness it suffices to check the fact that the following theorem*

(1) f *is locally of finite type. Since* $S = \mathrm{Spec}(R)$ *and* $Y = \mathrm{Spec}(R)$.

Proof. This is form all sheaves of sheaves on X. But given a scheme U and a surjective étale morphism $U \to X$. Let $U \cap U = \coprod_{i=1,\ldots,n} U_i$ be the scheme X over S at the schemes $X_i \to X$ and $U = \lim_i X_i$. □

The following lemma surjective restrocomposes of this implies that $\mathcal{F}_{x_0} = \mathcal{F}_{x_0} = \mathcal{F}_{\mathcal{X},\ldots,0}$.

Lemma 0.2. *Let* X *be a locally Noetherian scheme over* S, $E = \mathcal{F}_{X/S}$. *Set* $\mathcal{I} = \mathcal{J}_1 \subset \mathcal{I}'_n$. *Since* $\mathcal{I}^n \subset \mathcal{I}^n$ *are nonzero over* $i_0 \leq \mathfrak{p}$ *is a subset of* $\mathcal{J}_{n,0} \circ \overline{A}_2$ *works.*

Lemma 0.3. *In Situation* **??**. *Hence we may assume* $\mathfrak{q}' = 0$.

Proof. We will use the property we see that $\mathfrak{p}$ is the mext functor (**??**). On the other hand, by Lemma **??** we see that

$$D(\mathcal{O}_{X'}) = \mathcal{O}_X(D)$$

where K is an F-algebra where δ_{n+1} is a scheme over S. □

Logistics

Course Information

- **Name:** Introduction to Deep Machine Learning (ELEC 677)
- **Credits:** 3 hours
- **Location/Time:** Tuesdays 2:30 - 5pm DH 1042 (new room if needed)
- **Teaching Assistant:** Tan Nguyen <mn15@rice.edu>
- **Course Website:** elec677.rice.edu + Piazza (Discussion Forum)
- **Office Hours:**
  - Ankit Patel: Wednesdays 9:30 - 10:30am CST @Duncan Hall 2050
  - Tan Nguyen: Fridays 4-5pm CST @Duncan Hall 2030 (or 2050)

The Mission

- **Observations about Deep Learning (DL)**

  1. It works. (Finally.)

  2. It has an enormous number of potential applications in a wide variety of fields, many of which are just beginning to see DL's influence.

  3. There is a steep learning curve at the beginning.

  4. You are young and agile. If you invest now, you will reap the benefits.

- **Main Goal of this Course:** To jumpstart your ability to use Deep Learning in <u>your</u> research.

The Mission

- **Main Goal of this Course:** To jumpstart your ability to use Deep Learning in your research.
  - Designed for students who want to start using DL in their research
  - Myriad applications of DL in many many fields
  - Less Theory, More Doing: This is not a math class (though we will cover some aspects of theory of DL near the end)

Prerequisites

- **Level:** This class is for **GRADUATE** students, ready to do research! (I will grant exceptions to highly advanced undergrads who are ready to do research. Please see me after class)
- **Programming:** Entire course will be in python, using numpy, theano, and/or TensorFlow. Willingness to learn the DL software ecosystem (Linux, packages, git, etc.)
- **Calculus:** Differentiation, chain rule
- **Linear Algebra:** Vectors, matrices, eigenvalues/vector, Singular Value Decomposition.
- **Probability and Statistics:** random variables, multivariate Gaussians (mean covariance), Bayes Rule, Law of Total Probability, Conditional probabilities/expectations
- **Machine Learning:** Principal Component Analysis, Factor Analysis, Gaussian Mixture Models, train/test splitting, cross-validation, preprocessing and working with datasets.
- **Optimization:** Cost functions, taking gradients, penalty terms, implementing in code

Grading

- **Assignments:** 3-4 for total of 70%
  - Late Policy: 2 late days to allocate, after that 25% penalty per day
- **Final Project + Presentation:** 30%
- **Grader(s):** TBD
- **Note:** initial scores on assignments will be scaled based on distribution of student performance.

Assignments

- **Worth:** 3-4 assignments for total of 70% of the grade, due Tuesday **BEFORE** class at 11am CST
- **Submissions:** TBD (e.g. gradescope)
- Late Policy:

Final Project

- Up to two people per project
- Can combine across multiple classes if relevant. Please e-mail instructors of both courses to discuss.
- **Optional (but highly encouraged):** Choose final project topic relevant to your personal research/field. Please e-mail to discuss.

Questions?