Preprocessing data
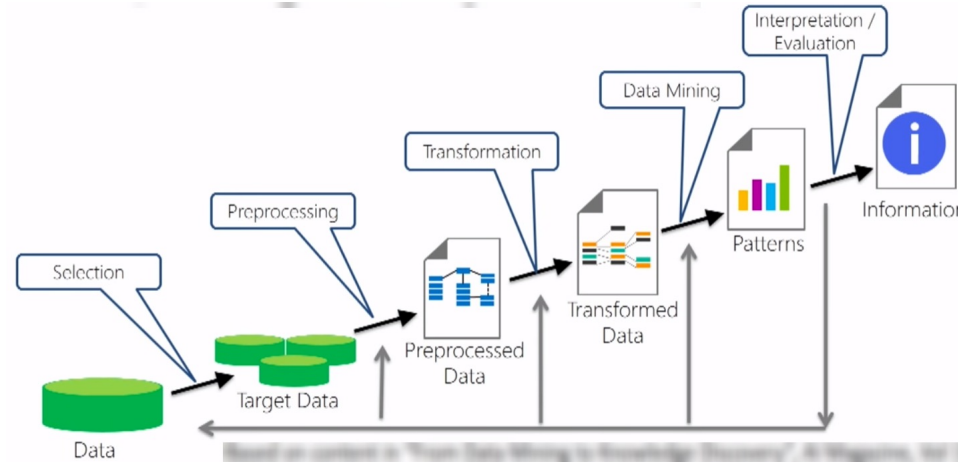
Prof. Roberto Fritsche-Neto

rfneto@agcenter.lsu.edu

Baton Rouge, April 3rd, 2024

KDD (Knowledge Discovery in Databases)

<u>Definition</u> : *"KDD is the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data"*
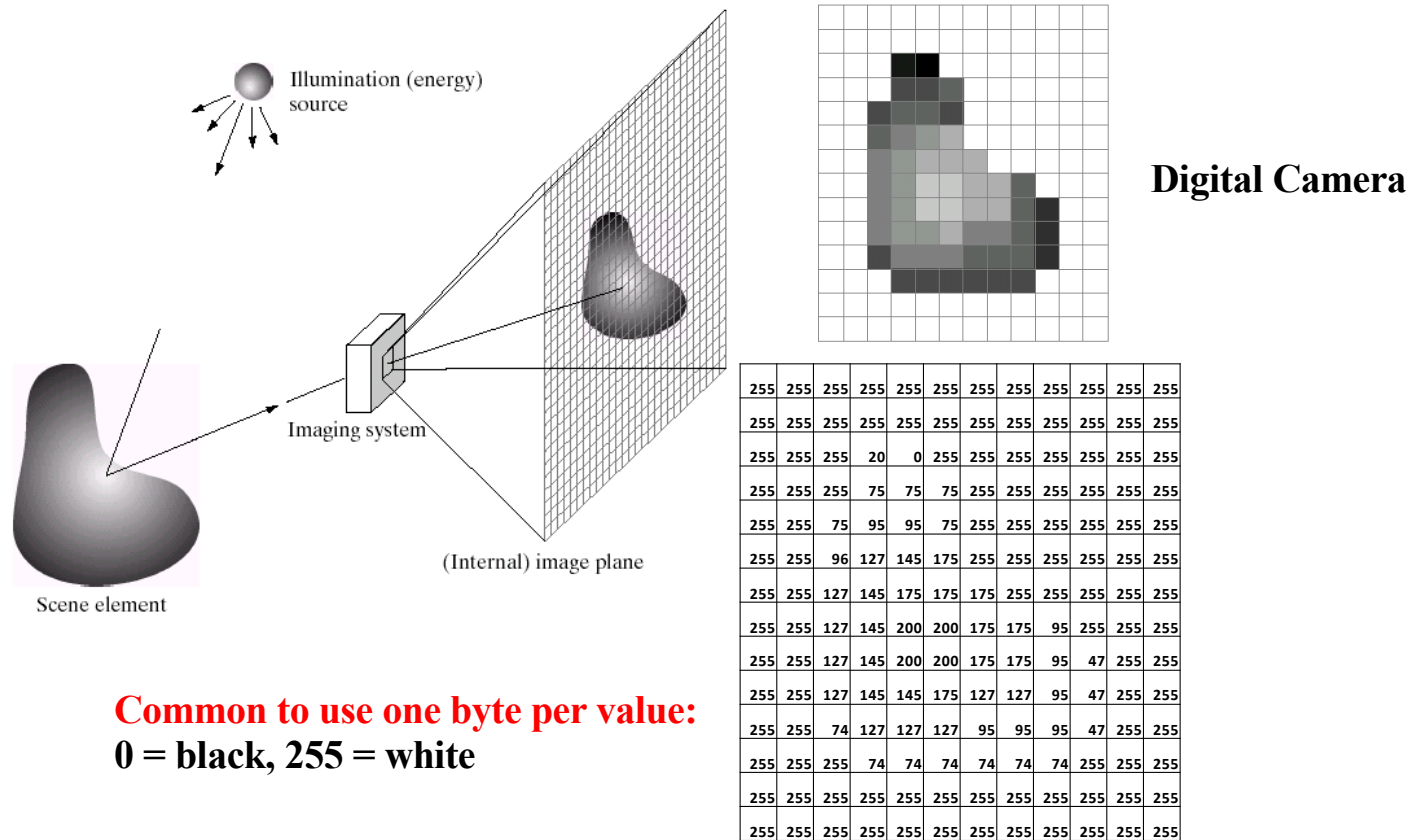

CNN Explainer

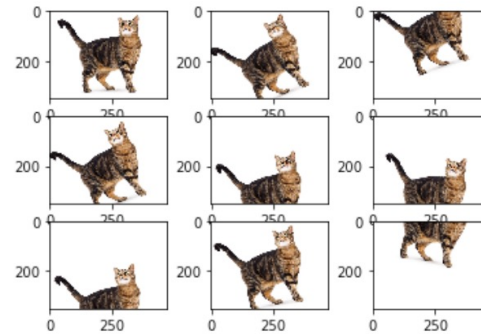Neural Network Playground

Interpretable machine learning

Extracting numbers from an image

Digital Camera

255	255	255	255	255	255	255	255	255	255	255	255
255	255	255	255	255	255	255	255	255	255	255	255
255	255	255	20	0	255	255	255	255	255	255	255
255	255	255	75	75	75	255	255	255	255	255	255
255	255	75	95	95	75	255	255	255	255	255	255
255	255	96	127	145	175	255	255	255	255	255	255
255	255	127	145	175	175	175	255	255	255	255	255
255	255	127	145	200	200	175	175	95	255	255	255
255	255	127	145	200	200	175	175	95	47	255	255
255	255	127	145	145	175	127	127	95	47	255	255
255	255	74	127	127	127	95	95	95	47	255	255
255	255	255	74	74	74	74	74	74	255	255	255
255	255	255	255	255	255	255	255	255	255	255	255
255	255	255	255	255	255	255	255	255	255	255	255

Common to use one byte per value:
0 = black, 255 = white
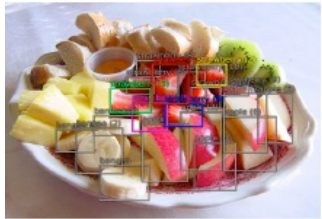
Convolutions kernels

Imagery augmentation

- **Artificial data generation in order to expand datasets as input for machine learning algorithms**
- **Mask, Elastic Distortions, perspectives, size, rotations, cropping, flip, and random erasing**
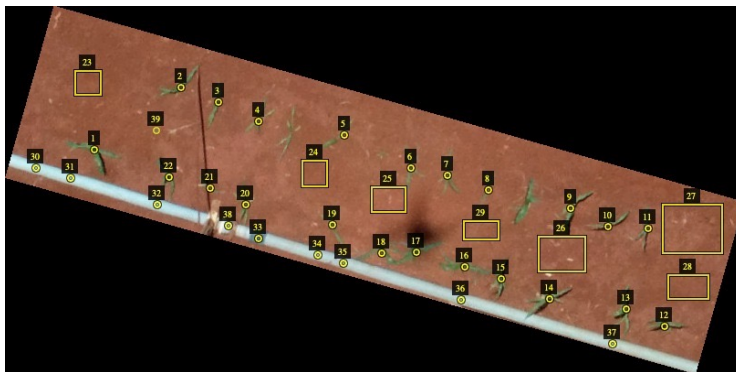

Image augmentation

Image annotation

Groundtruth:
strawberry
strawberry (2)
strawberry (3)
strawberry (4)
strawberry (5)
strawberry (6)
strawberry (7)
strawberry (8)
strawberry (9)
strawberry (10)
apple
apple (2)
apple (3)

- **Task of annotating an image with labels**

- **The first methods learned the correlations between image features and training annotations**

- **Type of multi-class image classification with a very large number of classes - as large as the vocabulary size**

- **The AI methods try to translate the textual vocabulary with the 'visual vocabulary', or clustered regions known as *blobs***

VGG tool

Image annotation

```
1 {"_via_settings":{"ui":{"annotation_editor_height":20,"annotation_editor_fontsize":0
  .9,"leftsidebar_width":18,"image_grid":{"img_height":80,"rshape_fill":"none","rshape_fill_opacity":0
  .3,"rshape_stroke":"yellow","rshape_stroke_width":2,"show_region_shape":true,"show_image_policy":"all"},"image":
  {"region_label":"__via_region_id__","region_color":"__via_default_region_color__","region_label_font":"10px
  Sans","on_image_annotation_editor_placement":"NEAR_REGION"},"core":{"buffer_size":18,"filepath":{},"default_filepath":""},"project":
  {"name":"LGN5838"}},"_via_img_metadata":{"parcelaRGB_356.png231022":{"filename":"parcelaRGB_356.png","size":231022,"regions":
  [{"shape_attributes":{"name":"point","cx":98,"cy":157},"region_attributes":{"maize":"1"}},{"shape_attributes":
  {"name":"point","cx":193,"cy":89},"region_attributes":{"maize":"1"}},{"shape_attributes":
  {"name":"point","cx":234,"cy":105},"region_attributes":{"maize":"1"}},{"shape_attributes":
  {"name":"point","cx":278,"cy":126},"region_attributes":{"maize":"1"}},{"shape_attributes":
  {"name":"point","cx":372,"cy":141},"region_attributes":{"maize":"1"}},{"shape_attributes":
  {"name":"point","cx":445,"cy":177},"region_attributes":{"maize":"1"}},{"shape_attributes":
  {"name":"point","cx":485,"cy":185},"region_attributes":{"maize":"1"}},{"shape_attributes":
  {"name":"point","cx":530,"cy":201},"region_attributes":{"maize":"1"}},{"shape_attributes":
  {"name":"point","cx":620,"cy":221},"region_attributes":{"maize":"1"}},{"shape_attributes":
  {"name":"point","cx":661,"cy":241},"region_attributes":{"maize":"1"}},{"shape_attributes":
  {"name":"point","cx":705,"cy":243},"region_attributes":{"maize":"1"}},{"shape_attributes":
  {"name":"point","cx":723,"cy":350},"region_attributes":{"maize":"1"}},{"shape_attributes":
  {"name":"point","cx":681,"cy":331},"region_attributes":{"maize":"1"}},{"shape_attributes":
  {"name":"point","cx":597,"cy":320},"region_attributes":{"maize":"1"}},{"shape_attributes":
  {"name":"point","cx":544,"cy":298},"region_attributes":{"maize":"1"}},{"shape_attributes":
  {"name":"point","cx":504,"cy":285},"region_attributes":{"maize":"1"}},{"shape_attributes":
  {"name":"point","cx":451,"cy":269},"region_attributes":{"maize":"1"}},{"shape_attributes":
  {"name":"point","cx":413,"cy":270},"region_attributes":{"maize":"1"}},{"shape_attributes":
  {"name":"point","cx":359,"cy":239},"region_attributes":{"maize":"1"}},{"shape_attributes":
  {"name":"point","cx":264,"cy":217},"region_attributes":{"maize":"1"}},{"shape_attributes":
  {"name":"point","cx":225,"cy":199},"region_attributes":{"maize":"1"}},{"shape_attributes":
  {"name":"point","cx":180,"cy":187},"region_attributes":{"maize":"1"}},{"shape_attributes":
  {"name":"rect","x":77,"y":71,"width":28,"height":26},"region_attributes":{"pipe":"","soil":"1","maize":""}},{"shape_attributes":
  {"name":"rect","x":326,"y":169,"width":27,"height":28},"region_attributes":{"pipe":"","soil":"1","maize":""}},{"shape_attributes":
  {"name":"rect","x":402,"y":198,"width":37,"height":27},"region_attributes":{"pipe":"","soil":"1","maize":""}},{"shape_attributes":
  {"name":"rect","x":585,"y":252,"width":51,"height":38},"region_attributes":{"pipe":"","soil":"1","maize":""}},{"shape_attributes":
  {"name":"rect","x":721,"y":217,"width":65,"height":53},"region_attributes":{"pipe":"","soil":"1","maize":""}},{"shape_attributes":
  {"name":"rect","x":727,"y":294,"width":44,"height":26},"region_attributes":{"pipe":"","soil":"1","maize":""}},{"shape_attributes":
  {"name":"rect","x":503,"y":235,"width":38,"height":20},"region_attributes":{"pipe":"","soil":"1","maize":""}},{"shape_attributes":
  {"name":"point","cx":34,"cy":177},"region_attributes":{"pipe":"1"}},{"shape_attributes":
  {"name":"point","cx":72,"cy":188},"region_attributes":{"pipe":"1"}},{"shape_attributes":
```