Anole classifier

1. Subject Line:

Progress on Machine Learning Model for Species Classification and Testing Data Refinement

2. Two-line Summary:

The team discussed the ongoing development of a machine learning model for classifying species from images, with near-perfect accuracy on training data but issues due to incorrect data splitting. Jacob Dallaire plans to retrain the model and leverage a new GPU to speed up the process.

3. Key Points & Repeated Topics:

- Jacob Dallaire trained a machine learning model using transfer learning and MobilenetV2 with ~2.5 million parameters, of which 2 million are trainable.
- The model had near-100% accuracy with training data, but there was an error in the data split for testing.
- Approximately 113,000 licensed open images across 5 species were used in training.
- The goal is to check accuracy for unseen data and split metrics (e.g., accuracy, recall, and precision) by species.
- There's a new GPU coming soon, which could reduce training time (currently ~20 hours).
- Future presentations will include example images and classification issues for review.
- Discussion of using one-hot encoding for predictions and reviewing bad classifications by checking the top two class choices.

4. Company / Project Mentions:

Company / Project Name	Description	Associated Company / Contact
MobilenetV2	Pre-trained model used as a base for transfer learning	N/A
Georgia Tech	Institution managing the IT department setting up the GPU	N/A

5. People Mentioned:

Name	Description	Company / Associated Entity
Jacob Dallaire	Machine learning specialist working on species classification	N/A
James T Stroud	Participant providing feedback on model progress	N/A
Jonathan J Suh	Person involved with the GPU setup	Georgia Tech
John	Likely part of IT or operations, mentioned by James Stroud	Georgia Tech (presumed)

6. Numbered Metrics / Data Points:

1. **2.5 million parameters** in the machine learning model, with **2 million trainable**.
2. **20 hours** to retrain the model.
3. **113,000 images** used for training, licensed as open.
4. Training data consists of **5 species**.
5. New GPU to be set up within **1-2 weeks**.

7. Action Items:

Responsible: Jacob Dallaire

- Retrain the model with a correct training/testing data split.
- Test model on unseen data and refine metrics by species (accuracy, recall, precision).
- Prepare future presentation with sample images showing correctly classified and misclassified images.
- Research how to shift the code from CPU to GPU to speed up processing.

Responsible: Jonathan J Suh / Georgia Tech IT

- Set up the new GPU, expected by Thursday or Friday.

Responsible: James T Stroud

- Review next updates and images once they are presented.

Xray

1. Subject Line:

Improvements in X-ray Landmark Detection and Error Reduction with Doctor Porto's Support

2. Two-line Summary:

The team discussed updates on X-ray image processing for species landmark detection, highlighting improvements from correcting bounding box placement and error reduction. Further discussions centered around visualizations, challenges with outliers, and future refinements to improve model accuracy on extremities.

3. Key Points & Repeated Topics:

- **X-ray Image Processing**: Improvements in bounding box placement, guided by advice from Dr. Porto, significantly reduced model errors.
- **Better Landmark Detection**: Use of bounding boxes for the entire X-ray image improved accuracy, particularly in detecting landmarks.
- **Dr. Porto's Contributions**: Assisted with code for bounding box placement, XML file modification, and building an inference script.
- **Visualization**: Visual comparison showed ground truth landmarks (green) closely matching model-predicted landmarks (red), with a focus on toe and fingertip errors.
- **Error Metrics**: Errors reduced to **below 1 millimeter** on average across 232 images, except for extremities like toes and fingertips.
- **Training Data Size**: Current dataset contains 232 images, with plans to expand it for better generalization.
- **Outlier Analysis**: Mercedes identified specific images with consistent outlier classification and suggested refining training data to improve model accuracy.
- **Importance of Body Alignment**: Straight body and limb positioning in X-rays help improve the model's classification accuracy.
- **Next Steps**: Plans to retrain the model with a focus on increasing data consistency, refining limb positioning, and comparing automated vs manual landmarking accuracy.

4. Company / Project Mentions:

Company / Project Name	Description	Associated Company / Contact
Dr. Porto's Project	Improvement of X-ray model accuracy by assisting with bounding box placement and inference script.	N/A
Georgia Tech IT	Handles setup of computing resources, including GPU setup.	N/A

5. People Mentioned:

Name	Description	Company / Associated Entity
Mercedes Quintana	Lead on visualizations and data analysis, focused on reducing outlier impact.	N/A
Ayush Parikh	Provided model updates and presented results of bounding box changes for X-ray images.	N/A
James T Stroud	Provides feedback and direction on project, asks for visual comparisons of predictions.	N/A
Jonathan J Suh	Works on X-ray image preprocessing, landmark identification, and collaborates with Mercedes.	Georgia Tech
Dr. Porto	Provided expert advice on improving the X-ray model, including bounding box placement.	N/A
John	Assisted with landmarking X-ray images and ensuring data consistency.	N/A

6. Numbered Metrics / Data Points:

1. **232 images** in the current automatically processed dataset.
2. Errors reduced to **below 1 millimeter** across most landmarks, except for toes and fingertips.
3. Outlier image **#8** had **17 outliers** across **34 landmarks**.
4. Image **#10** had **0 outliers**, classified as well-behaved.
5. **57-59 images** were analyzed for outlier classification.
6. Average model errors now consistently below **1.5 standard deviations** for most images.

7. Action Items:

Responsible: Mercedes Quintana

- Review visualizations for 57-59 images and check discrepancies in data for the missing two images.
- Analyze and filter out problematic images with inconsistent body alignment for better training results.
- Compare shape analysis on extremities and refine the model to improve toe and fingertip accuracy.
- Investigate grid search to explore model parameters, including depth, to enhance performance.

Responsible: Ayush Parikh

- Retrain the model to incorporate Dr. Porto's new bounding box code and explore increasing the training dataset size.
- Compare automatic landmarks with manually processed images to validate model accuracy.

Responsible: Jonathan J Suh

- Verify landmark consistency in processed X-ray images, particularly in the jawline and eye socket areas.
- Assist with manual relabeling of images to ensure consistency across landmarks.

Responsible: James T Stroud

- Provide feedback on visual outputs, particularly on extremities like fingertips and toes.
- Facilitate further analysis and comparison between automated and manually processed images.

Jaws

1. Subject Line:

Progress on Lizard Jaw Segmentation and Future Pipeline Development for Automated CT Scan Analysis

2. Two-line Summary:

The team discussed the ongoing development of a lizard jaw segmentation pipeline using a state-of-the-art mask segmentation model. The goal is to automate CT scan processing for future comparative analysis, reducing manual work and integrating the pipeline for broader scientific and community use.

3. Key Points & Repeated Topics:

- **Lizard Jaw Segmentation Project**: Philip G. Woolley updated on progress using a segmentation model on 2D sagittal slices of CT scans.
- **Segmentation Model**: A mask segmentation transformer model was trained on **235 manually segmented slices**.
- **Preliminary Results**: The model outputs are promising, although training data is currently limited. Manual segmentations are used for comparison.
- **Model Output**: The model segments the jaw and teeth and labels other structures, with good performance considering the limited training data.
- **Next Steps**:
    - **Segment more training data** to improve model accuracy.
    - Evaluate model performance on a full set of scans.
    - Develop a method to combine 2D slice segmentations into a 3D volume.
    - Begin steps 1 and 2 of the pipeline, which involve cleaning and preparing original lizard CT scans for segmentation.
- **Importance of Pipeline**: James T. Stroud emphasized the need for the pipeline to automate future CT scan processing for comparative analysis across species, reducing manual time for handling large datasets.

4. Company / Project Mentions:

Company / Project Name	Description	Associated Company / Contact
Lizard Jaw Segmentation	Project to segment lizard jaws and teeth from CT scan slices	N/A
Mask Segmentation Transformer	Model used for automated segmentation of CT scan slices	N/A
Slicer	Software used for manual segmentation and volume reconstruction	N/A

5. People Mentioned:

Name	Description	Company / Associated Entity
Philip G. Woolley	Lead on the lizard jaw segmentation project, providing updates on progress	N/A
James T. Stroud	Provides feedback and direction, emphasizes the importance of automation	N/A

6. Numbered Metrics / Data Points:

1. **235 slices** used for training the segmentation model.
2. The project aims to process **5000 lizards**, including both live and museum specimens.
3. The model segments **upper jaw, lower jaw**, and **teeth** into distinct classes.

7. Action Items:

Responsible: Philip G. Woolley

- Segment additional training data to improve model accuracy.
- Test the model on a full scan to evaluate performance across multiple slices.
- Develop a method to stitch 2D segmentations into a usable 3D volume in Slicer.
- Begin work on steps 1 and 2 of the pipeline (initial image cleaning and head segmentation).

Responsible: James T. Stroud

- Use the developed pipeline to perform comparative analysis between species on existing and future CT scans.
- Continue supporting the development of the pipeline for broader applications, including citizen science integration.