AUTOMATED SCENARIO DESCRIPTION FOR IMAGES

Satish P
Siddharth G
Vijayalakshimi MLS

R S Milton (Supervisor)

April 11th, 2016

Problem Statement

To automatically identify creatures called Pokemon from cartoon images, the attacks which they execute, and generate natural language sentences from these realizations.

What Our System Does

Figure: Razor Leaf is demonstrated by Bulbasaur

Architectural Design of the Proposed System

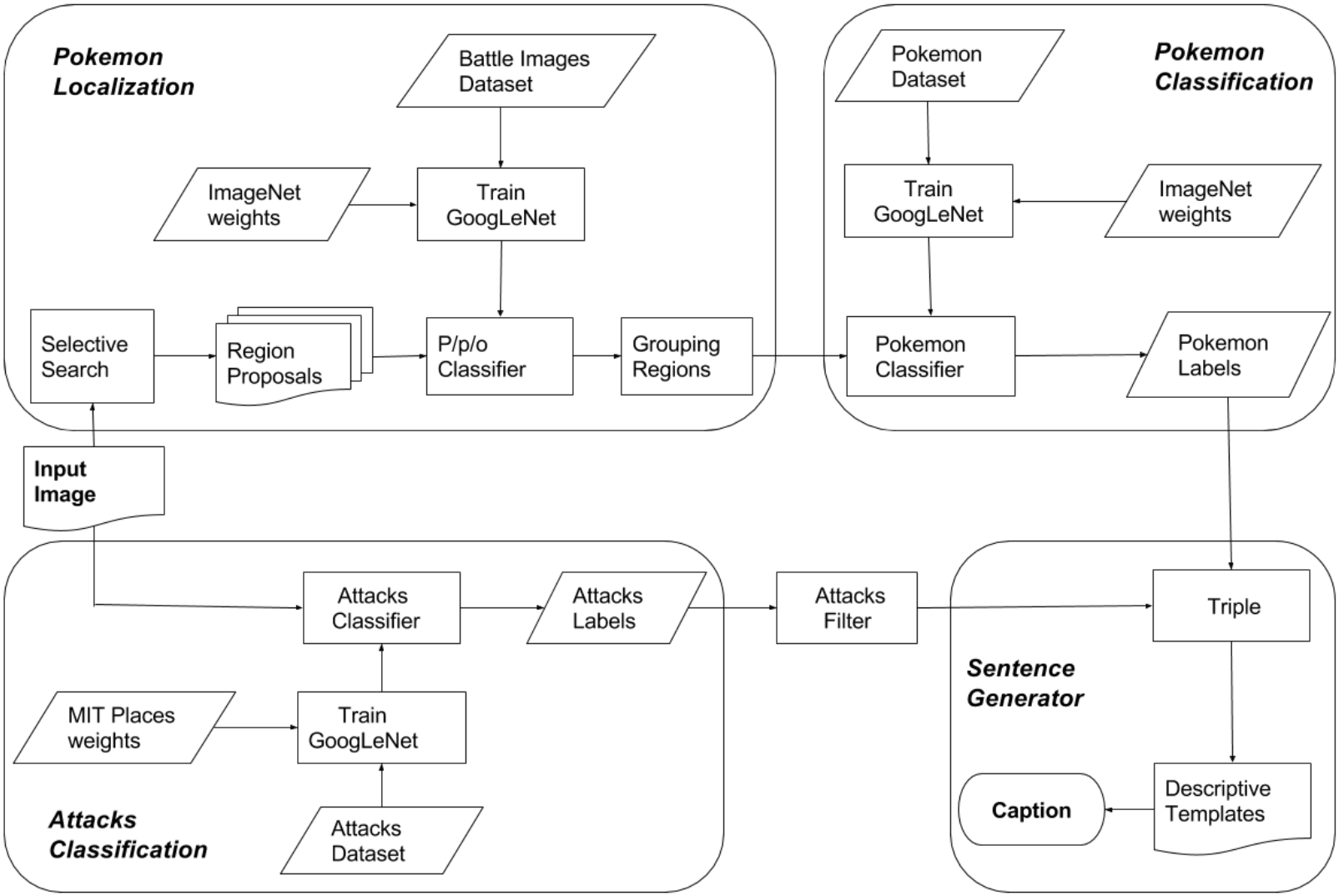


Figure: Achitecture of the image description pipeline

Dataset Formulation and Summary

Dataset	Classes	Train		Test		Total no. of images
		Images per class	No. of images	Images per class	No. of images	
Raw Vectorized Dataset	150	9	1349	4	600	1949
Backgrounded Dataset	150	135	20209	54	8100	28309
Augmented Dataset	150	33	4950	14	2100	7050
Backgrounded and Augmented Dataset	150	628	94137	269	40371	134508
Battle Images Dataset	3	Full - 210 Part - 1666 Other - 925	2801	Full - 62 Part - 499 Other - 277	838	3639
Augmented Attacks Dataset	144	103	14832	44	6356	21188

Selective Search: Object Localization

Object detection algorithm which proposes various regions in an image that are highly probable to contain an object in them.

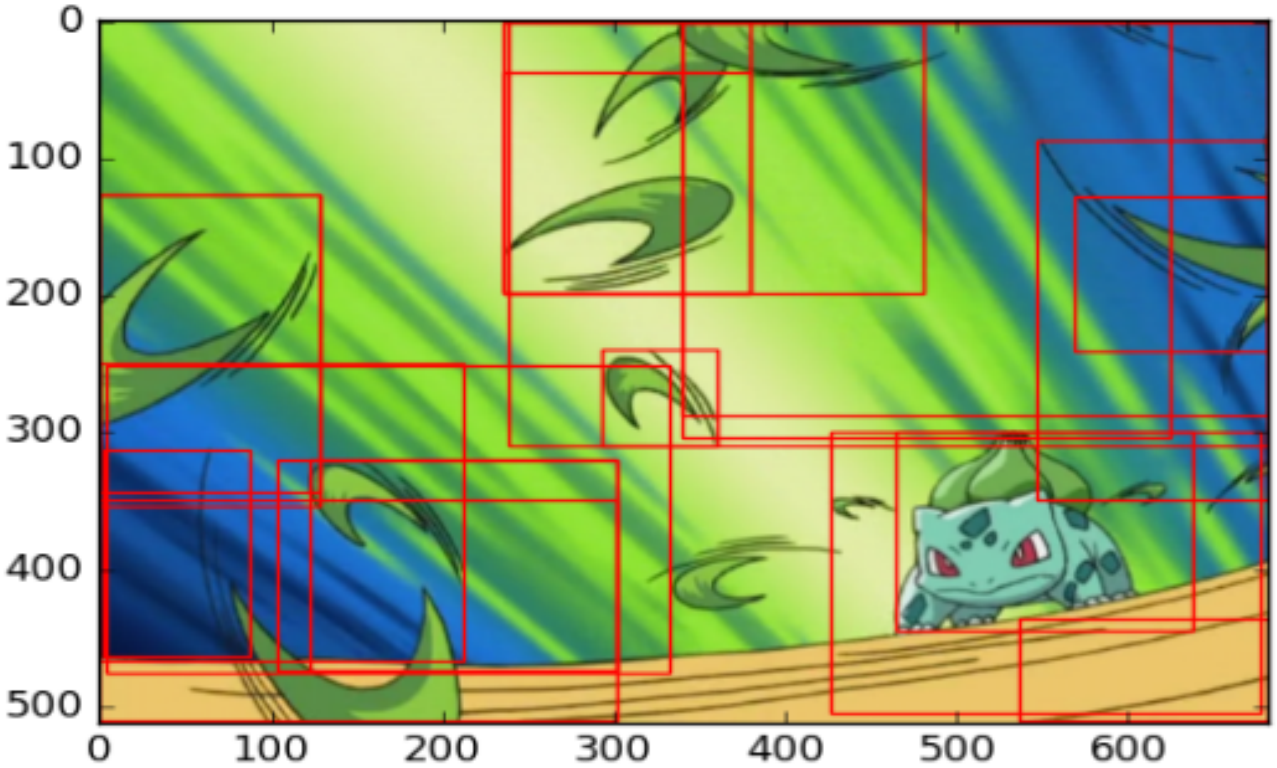
Fine Tuned Parameters

- Scale - 350, 450, 500
- Sigma - 0.8
- Min_Size - 30, 60, 120
- Min_Area - 2000

Customized Grouping Mechanisms

- Merge Concentric Proposals, Contained Box Removal, Draw Super Box

Selective Search: Example



Pokemon/Parts/Others Classifier

- Takes previously cropped regions of battle image as input
- Classifies into one of three categories - Pokemon/Parts/Other
- GoogLeNet architecture - Convolutional Neural Network
- Initialized with ImageNet weights
- Accuracy achieved: 74.9%

Pokemon/Parts/Others Classifier: Output

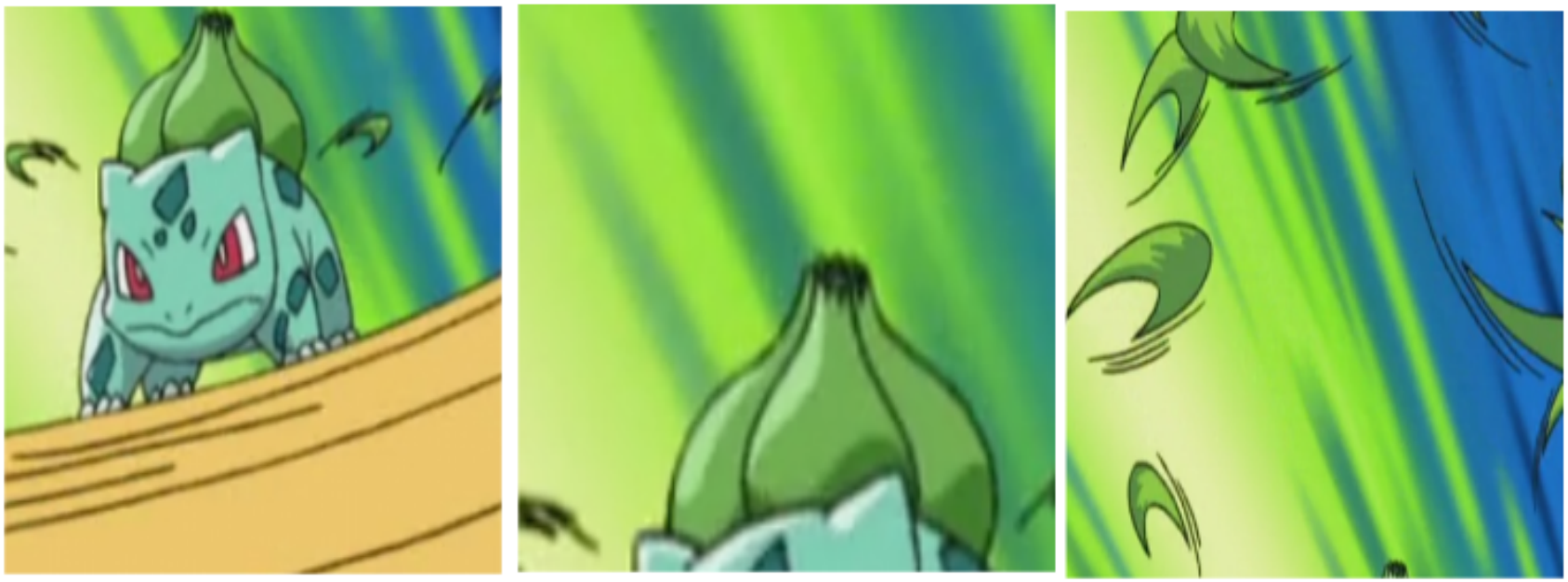


Figure: Pokemon, Parts and Others (L to R)

Pokemon Classification

Convolutional Neural Networks: Deep Neural Networks built specifically to understand images.

- We train GoogLeNet again - a CNN
- Initialized with ImageNet weights

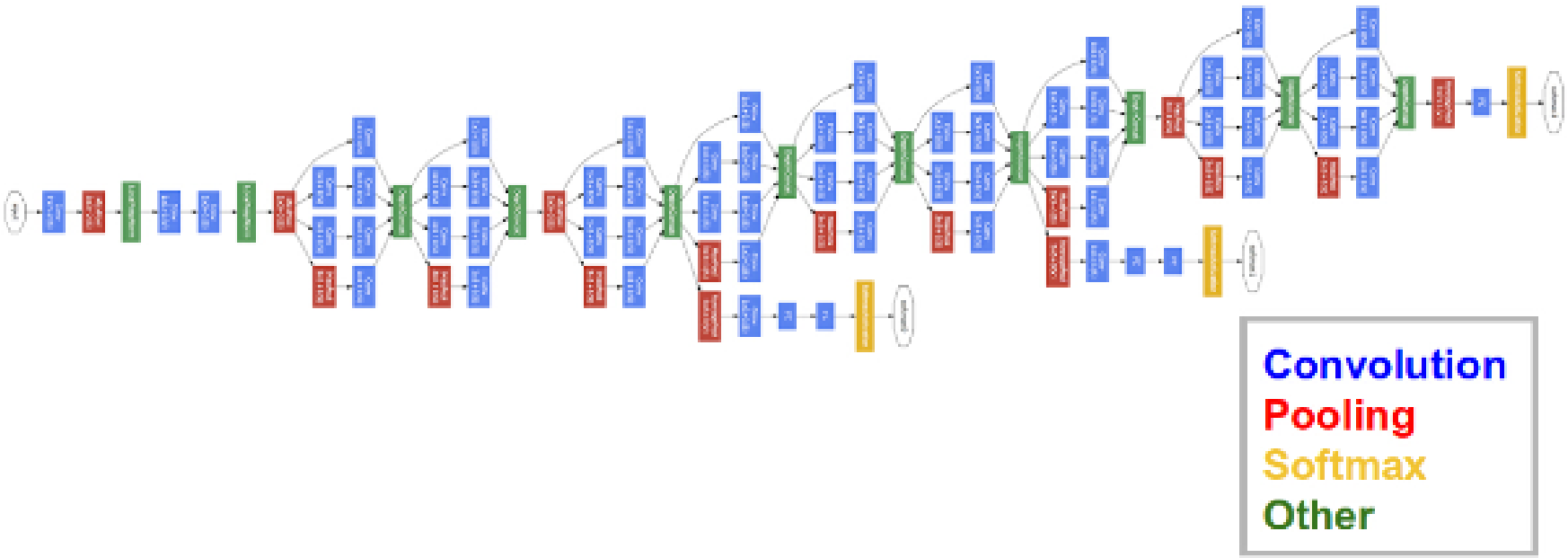


Figure: GoogLeNet Architecture

Pokemon Classification: SGD Solver

```
net: "googlenet/train_val.prototxt"
test_iter: 1407
test_interval: 4000
test_initialization: false
display: 40
base_lr: 0.001
lr_policy: "step"
stepsize: 32000
gamma: 0.1
max_iter: 10000000
momentum: 0.9
weight_decay: 0.0002
snapshot: 4000
snapshot_prefix: "bvlc_googlenet_pokenet"
solver_mode: GPU
```

Pokemon Classification: Results Obtained

Table: Results obtained with CNN

Levels of testing in the GoogleNet	Augmented Dataset		Backgrounded and Augmented Dataset	
	Top1	Top5	Top1	Top5
Level 1 (1/3 rd of Network)	74.39%	91.80%	79.22%	93.99%
Level 2 (2/3 rd of Network)	76.28%	92.30%	82.06%	95.32%
Level 3 (Full Network)	85.39%	96.43%	91.16%	98.03%

Pokemon Classification: Output

Figure: Top 5 guesses - Bulbasaur, Jynx, Arbok, Lapras, Machop

Attacks Classification

Consider attacks to be scenes rather than localized segments in an image. Attacks may not have screen presence, like “tackle”, and hence the whole image should be taken for classification.

- Train GoogLeNet
- Initialized with MIT Places 205 weights
- Accuracy Obtained: 56%

Attacks Classification: Output

Figure: Top 5 guesses - Razor Leaf, Leech Seed, Vine Whip, Tackle, Sleep Powder

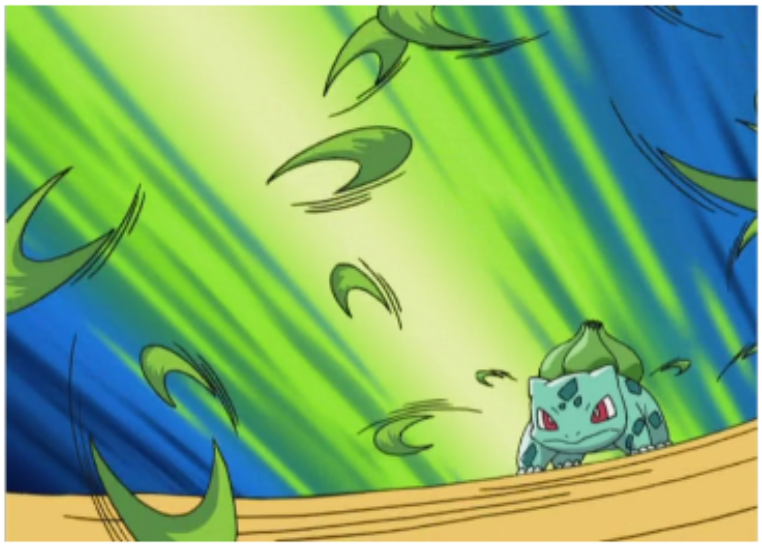
Filtering Attacks

Each Pokemon is capable of executing only a small number of attacks, and hence we formed a database of Pokemon-Attacks mapping, and used this in removing attacks which cannot be executed by the Pokemon identified in the image.

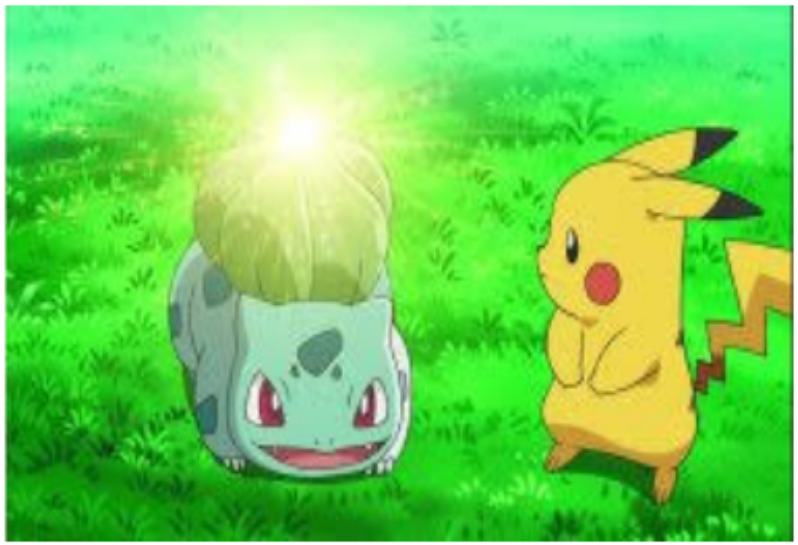
- **Pikachu** - ThunderBolt, ThunderWave, VoltTackle, IronTail, Tackle, Agility, Thunder, TailWhip, Growl, ElectroBall
- **Charmander** - FlameThrower, Tackle, Ember, TailWhip, Growl, Scratch, SmokeScreen, FlameBurst, FireSpin, Inferno
- **Squirtle** - WaterGun, Withdraw, Tackle, TailWhip, WaterPulse, HydroPump, IronDefense, SkullBash, Bubble, AquaTail

Triple Generation: Example

Triple Template : <PokemonA, PokemonB, AttackC>



<Bulbasaur, _____ , Razor Leaf >

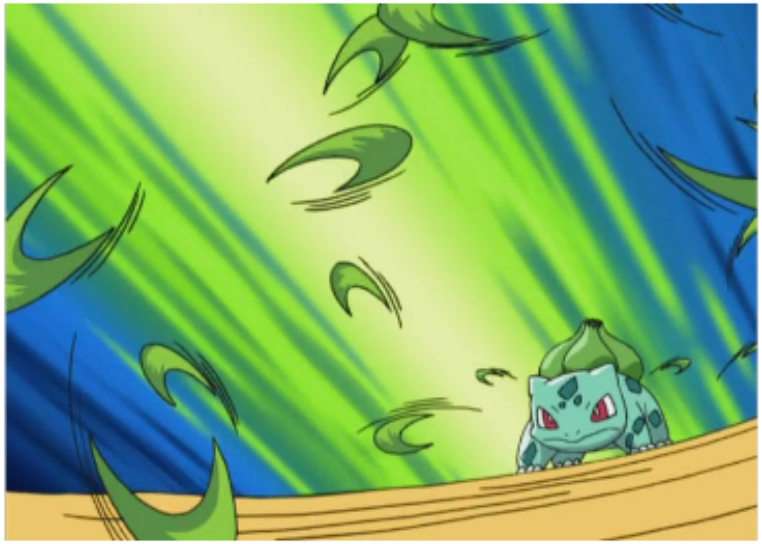


<Bulbasaur, Pikachu, Growth>

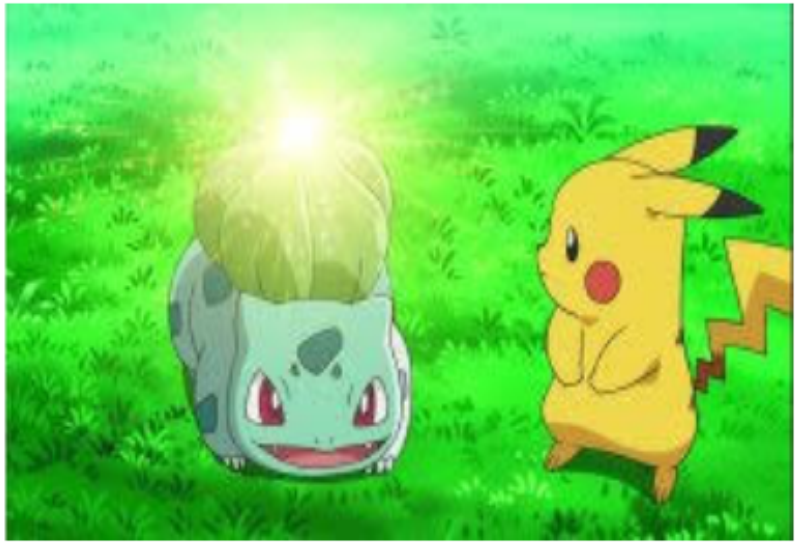
Sentence Formation

```
$attackB was used by $pokemonA
$pokemonA avails $attackB
$pokemonA availed $attackB
$pokemonA is availing $attackB
$pokemonA has availed $attackB
$attackB is availed by $pokemonA
$attackB is being availed by $pokemonA
$attackB has been availed by $pokemonA
$attackB was availed by $pokemonA
$pokemonA utilizes $attackB
$pokemonA utilized $attackB
$pokemonA is utilizing $attackB
$pokemonA has utilized $attackB
$attackB is utilized by $pokemonA
$pokemonA employed $attackC on $pokemonB
$pokemonA is employing $attackC on $pokemonB
$pokemonB is attacked by $pokemonA by employing $attackC
$pokemonB was attacked by $pokemonA by employing $attackC
$pokemonB is being attacked by $pokemonA by employing $attackC
$attackC is employed by $pokemonA to attack $pokemonB
$attackC was employed by $pokemonA to attack $pokemonB
$attackC is being employed by $pokemonA to attack $pokemonB
$pokemonA has attacked $pokemonB by employing $attackC
$pokemonA has employed $attackC on $pokemonB
$pokemonB has been attacked by $pokemonA by employing $attackC
```

Sentence Formation: Example



Bulbasaur deployed Razor Leaf



Bulbasaur uses growth on Pikachu

References

1. Deng, J. and Dong, W. and Socher, R. and Li, L.-J. and Li, K. and FeiFei, L. (2009) *ImageNet: A large-scale hierarchical image database*, CVPR
2. Jia, Yangqing and Shelhamer, Evan and Donahue, Jeff and Karayev, Sergey and Long, Jonathan and Girshick, Ross and Guadarrama, Sergio and Darrell, Trevor (2014), *Caffe: Convolutional Architecture for Fast Feature Embedding*, 22nd ACM international conference on Multimedia, arXiv:1408.5093, 675 - 678
3. Szegedy, Christian and Liu, Wei and Jia, Yangqing and Sermanet, Pierre and Reed, Scott Anguelov, Dragomir and Erhan, Dumitru and Vanhoucke, Vincent and Rabinovich, Andrew (2015) *Going Deeper With Convolutions*, CVPR, arXiv:1409.4842, 1-9
4. Uijlings, Jasper R. R. and van de Sande, Koen E. A. and Gevers, Theo and W. M. Smeulders, Arnold W. M (2013) *Selective Search for Object Recognition*, IJCV, Volume 104 (2), page 154-171
5. Zhou, B. and Lapedriza, A. and Xiao, J. and Torralba, A. and Oliva, A. (2014) *Learning Deep Features for Scene Recognition using Places Database*. Advances in Neural Information Processing Systems 27 (NIPS) spotlight
6. Pokemon - The Pokemon Company International, http://www.pokemon.com/us/
7. Bulbapedia - a community driven Pokémon encyclopedia, bulbapedia.bulbagarden.net
8. Pokemon Wiki , pokemon.wikia.com