How Deep Learning can Solve Phishing

1. Who Cares?
2. Research
3. Data
4. Challenges
5. Who should collaborate

Chris Cleveland, cc3580
Who Cares?
Who Cares?

- RSA Phish Report (2013 Review)
  - $12B/yr in losses to private sector
Who Cares?

• RSA Phish Report (2013 Review)
  – $12B/yr in losses to private sector
  – Global affect on business and government on every major continent: US, EMEA, LATAM, APJ
Who Cares?

• RSA Phish Report (2013 Review)
  – $12B/yr in losses to private sector
  – Global affect on business and government on every major continent: US, EMEA, LATAM, APJ

• McAfee: “Look at any high profile attack in the news and the initial vector of compromise is almost always a phishing email”
Research

• Old Inspiration
Research

- Old Inspiration
Research

• Old Inspiration
Research

• Old Inspiration

<table>
<thead>
<tr>
<th>PhishZoo approach: profile content</th>
<th>Accuracy</th>
<th>False positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>HTML</td>
<td>21.5%</td>
<td>1%</td>
</tr>
<tr>
<td>Visible text in HTML</td>
<td>70.3%</td>
<td>0.5%</td>
</tr>
<tr>
<td>Images</td>
<td>82.7%</td>
<td>2.5%</td>
</tr>
<tr>
<td>Image and visible texts</td>
<td>96.4%</td>
<td>1.4%</td>
</tr>
<tr>
<td>Screenshots</td>
<td>81.1%</td>
<td>30.3%</td>
</tr>
<tr>
<td>Keywords</td>
<td>97.6%</td>
<td>18.7%</td>
</tr>
<tr>
<td>Image and Keywords</td>
<td>90.2%</td>
<td>0.5%</td>
</tr>
</tbody>
</table>
Research

• Old Inspiration
• New State of the Art
  – Karen Simonyan, Andrew Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, arxiv 1409.1556, 2014
    • 1st place in localization (2014 ImageNet)
    • 2nd place in classification (2014 ImageNet)
Research

• Old Inspiration
• New State of the Art
    • Efficient multi-scale ConvNet implementation
Research

- Old Inspiration
- New State of the Art
  - NLP ???
    - Important for URL analysis and HTML content
Data

- PhishTank: ~1,500 validated reports daily
Data

• PhishTank: ~1,500 validated reports daily
Data

• PhishTank: \( \sim 1,500 \) validated reports daily
Data

• PhishTank :~1,500 validated reports daily

```
array(
    array(
        'phish_id' => 123456,
        'url' => 'http://www.example.com/',
        'submission_time' => '2009-06-19T15:47+00:00',
        'verified' => 'yes',
        'verification_time' => '2009-06-19T15:37:31+00:00',
        'online' => 'yes',
        'target' => '1st National Example Bank',
        'details' => array(
            array(
                'ip_address' => '1.2.3.4',
                'cidr_block' => '1.2.3.0/24',
                'announcing_network' => '1234',
                'rir' => 'arin',
                'detail_time' => '2006-10-01T02:30:54+00:00'
            )
        )
    )
)
```
Data

- PhishTank: ~1,500 validated reports daily
Data

• Synthesize:
  – Standard
    • Scale invariance
    • Translation invariance
Data

• Synthesize:
  – Standard
    • Scale invariance
    • Translation invariance
  – Color Inversion
Data

• Synthesize:
  – Standard
    • Scale invariance
    • Translation invariance
  – Color Inversion
  – Small geometric distortion
    • Rotation, affine
Data

- Synthesize:
  - Standard
    - Scale invariance
    - Translation invariance
  - Color Inversion
  - Small geometric distortion
    - Rotation, affine
  - Subtle effects (eg. light gradient)
Challenges

• Can we capture trustworthiness?
Challenges

• Can we capture trustworthiness?
Challenges

- Can we capture trustworthiness?
Challenges

• Can we capture trustworthiness?
• Variants
Challenges

• Can we capture trustworthiness?
• Variants
• Lopsided data
  – Phishing pages are a tiny subset of web pages
  – Within a particular page, trademark or ‘bait’ images like logos and brand iconography is a localized
Challenges

- Can we capture trustworthiness?
- Variants
- Lopsided data
- PhishTank data is tip of the iceberg
Challenges

• Can we capture trustworthiness?
• Variants
• Lopsided data
• PhishTank data is tip of the iceberg
Who should join (ideally)?

- Two collaborators (vision + NLP)
- Security background or interest is helpful but not required (big picture)
- Desire to make impact
- Ok with not solving full scope of problem within class but, again, passionate about making impact