

# Malware Detection

---

# 차례

---

- 목표
- 기존 연구의 문제점
- Dataset
- Library
- 시스템 구성도
- 성능

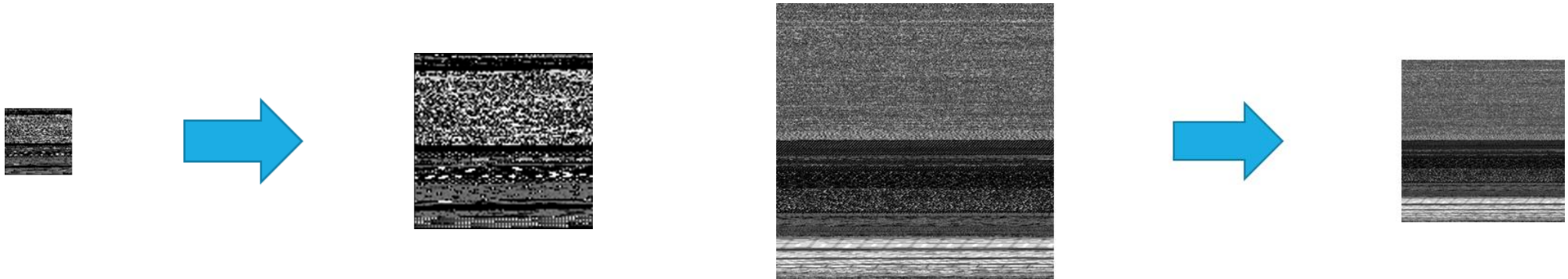
# 목표

---

- Malware 파일을 이미지화 시켜 malware 역할을 하는 부분을 detection 한다.
- Malware 파일을 이미지화 시켜서 convolution layer 거쳤을 때 activation되는 영역을 RCNN으로 학습시켜 malware 역할을 하는 부분을 detection 한다.
- 기존의 CNN을 이용해 malware를 판단하면 큰 크기의 malware의 특징 점들이 뭉게지기 때문에 큰 크기의 파일은 판단하기 어려울 수 있는 문제점이 있다.

# 기존 연구의 문제점

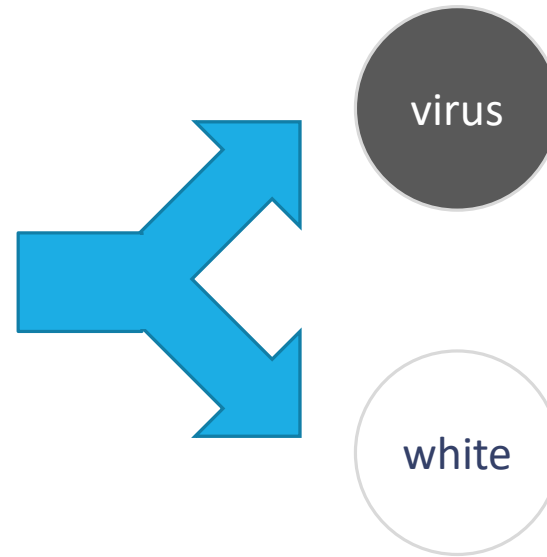
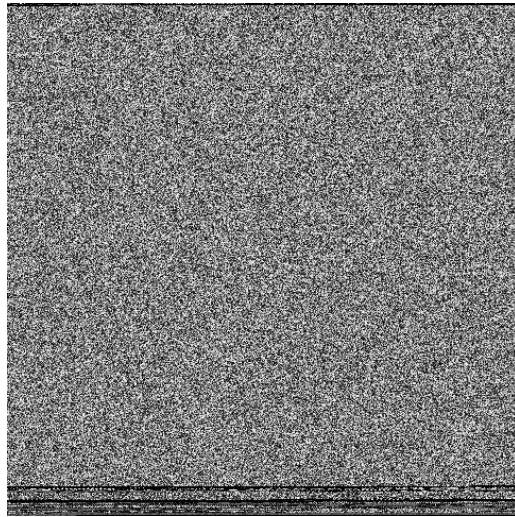
- CNN에 입력되는 이미지의 resize 예



- 기존의 CNN을 이용해 malware를 판단하면 큰 크기의 malware의 특징 점들이 묻게 지기 때문에 큰 크기의 파일은 판단하기 어려울 수 있는 문제점이 있다.

# 기존 연구의 문제점

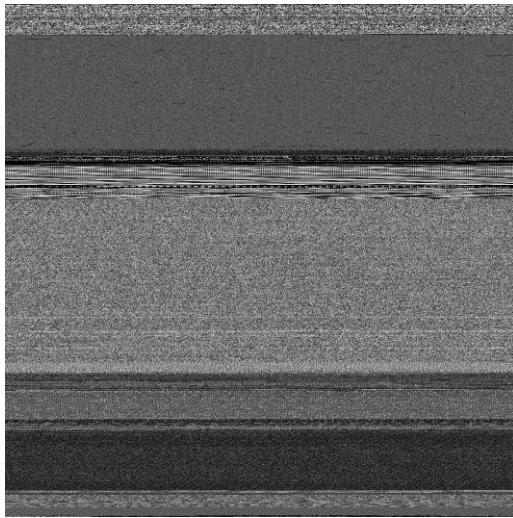
- CNN을 이용한 악성코드 판단의 예



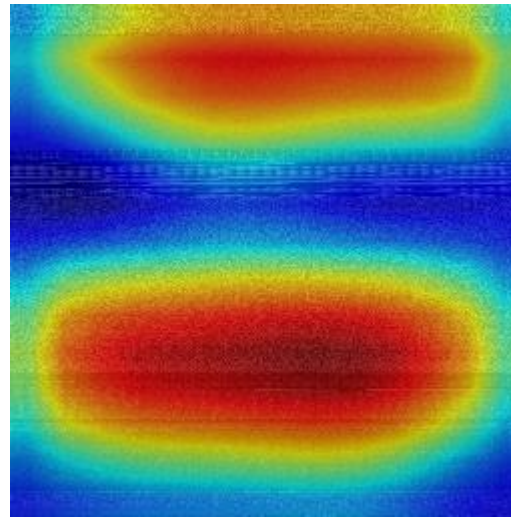
- CNN을 학습시켰을 때 activation 되는 영역을 이용하여 RCNN을 학습시켜서 큰 크기 파일에서 여러 Malware 영역을 detect한다.

# 기존 연구의 문제점 : 해결 방안

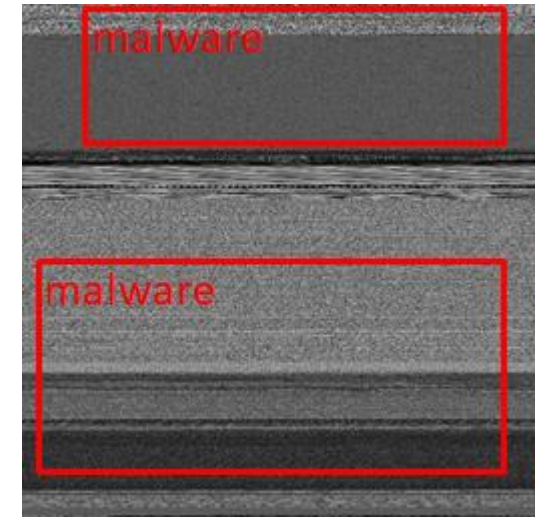
- 악성코드와 일반 파일을 이미지화한 파일을 CAM(Class Activation Mapping)을 이용해 활성화 영역을 학습하여 파일의 크기를 조정하지 않고 과 R-CNN을 통해 판단한다.
- 또한 CAM과 R-CNN을 이용하면 악성코드 동작 영역을 파악할 수 있다.



이진 파일 이미지



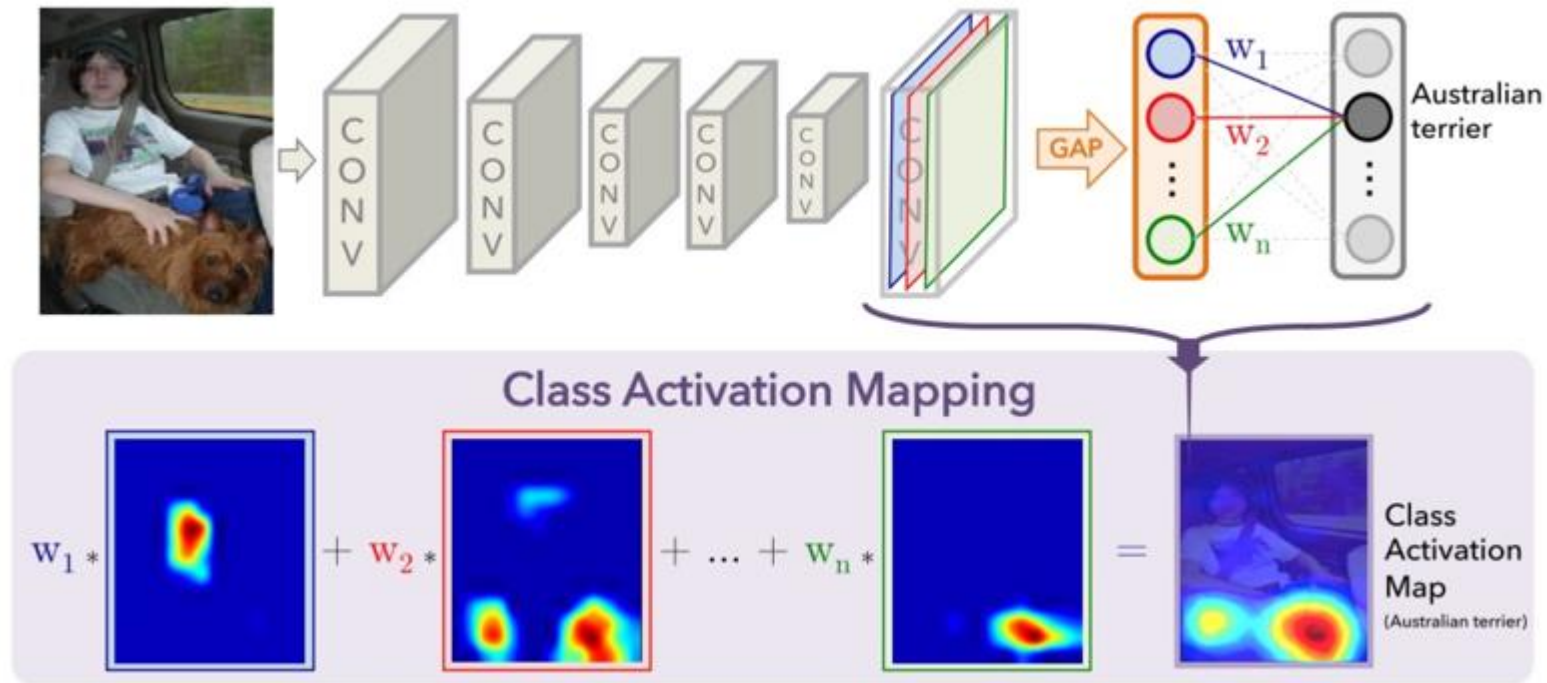
Class 활성화 영역



R-CNN 결과

# 기존 연구의 문제점 : 해결 방안

- CAM(Class Activation Mapping)

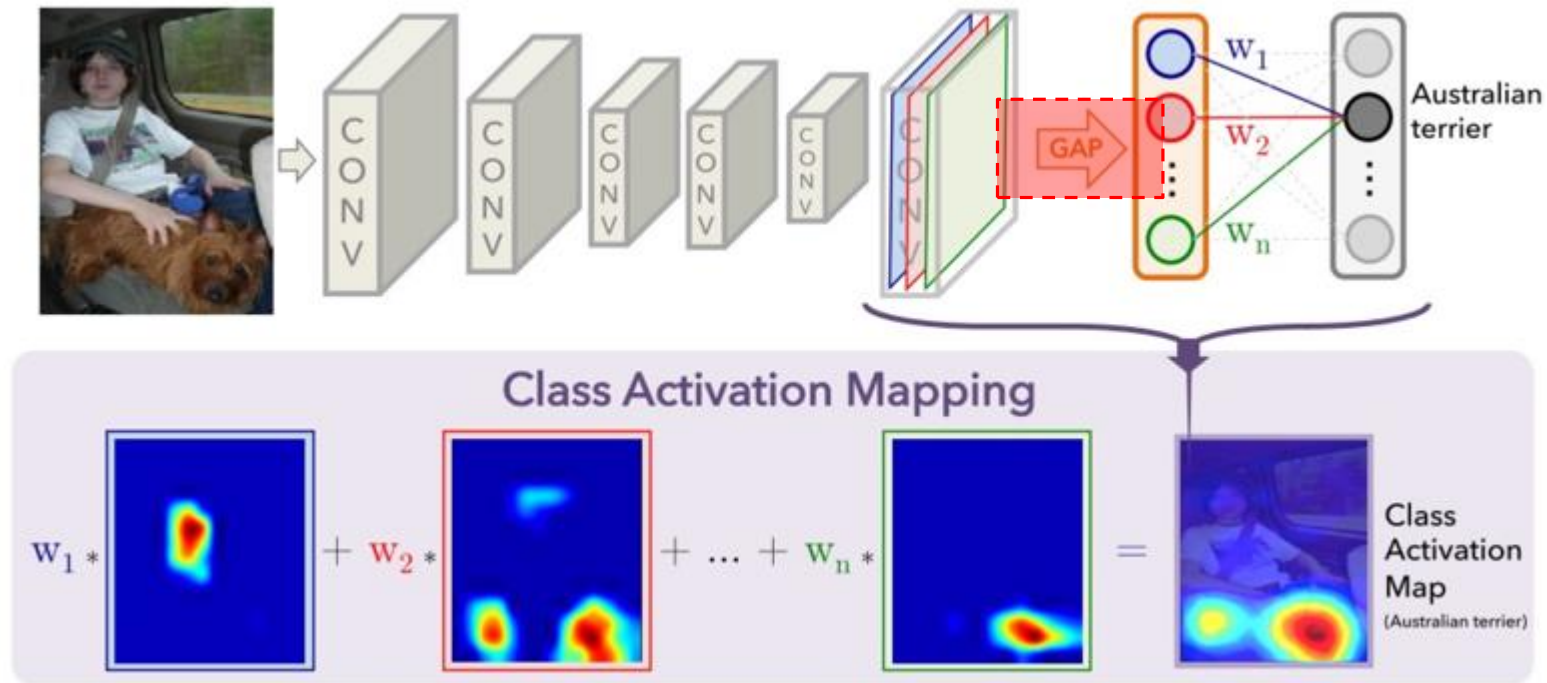


B. Zhou, A. Khosla et al., "Learning Deep Features for Discriminative Localization," CVPR 2016.



# 기존 연구의 문제점 : 해결 방안

- CAM(Class Activation Mapping)



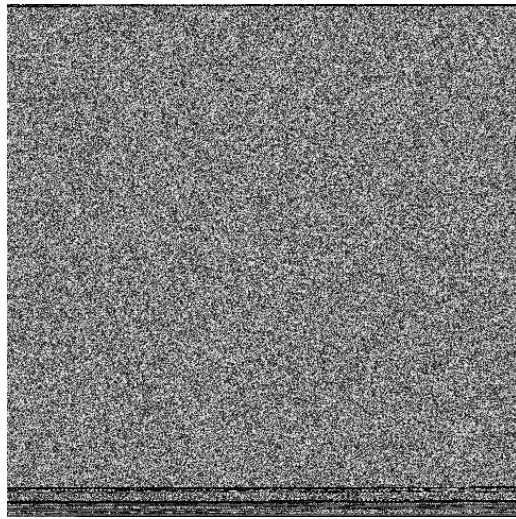
B. Zhou, A. Khosla et al., "Learning Deep Features for Discriminative Localization," CVPR 2016.



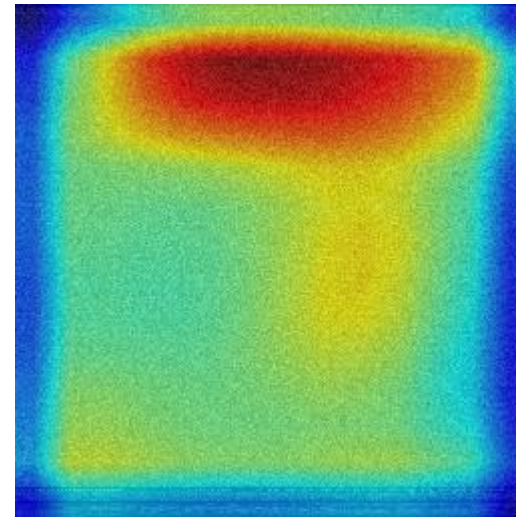
# 기존 연구의 문제점 : 해결 방안

---

- CAM(Class Activation Mapping)
  - CAM은 이미지가 FC layer를 지나면서 특정 class에 대해 활성화 되는 영역을 보여준다.



이진 파일 이미지

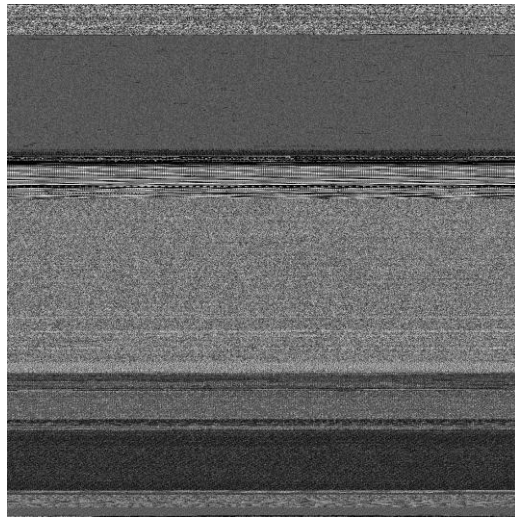


Class 활성화 영역

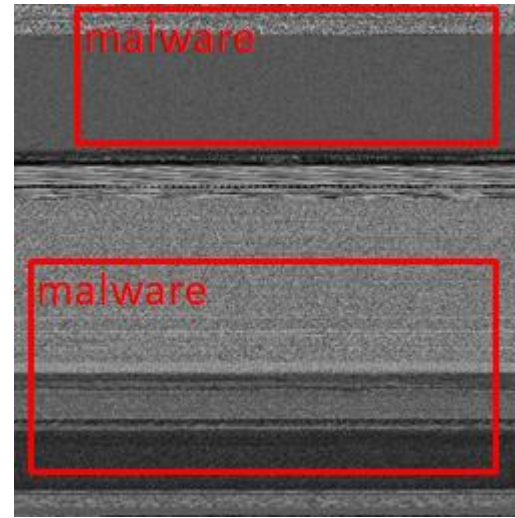
# 기존 연구의 문제점 : 해결 방안

---

- R-CNN
  - 이미지 전체에서 class에 해당하는 객체를 찾고 객체 영역을 출력한다.



이진 파일 이미지

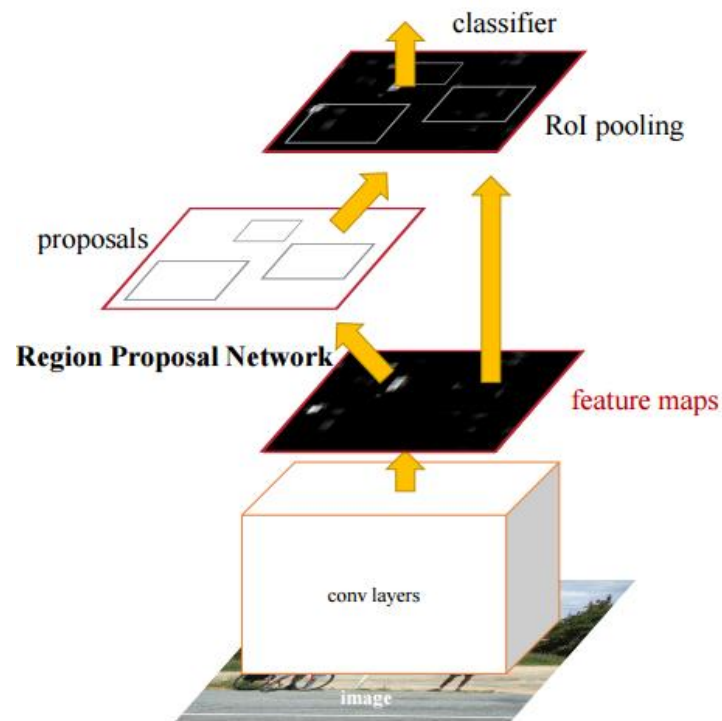


R-CNN 결과

# Library

---

- Faster R-CNN



Ren, Shaoqing, et al., “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” CVPR 2015.

# Dataset

---

- **Virus** - 20,462개
- **White** - 19,766개

**총 40,228개의 데이터로 구성**

# Dataset

---

- **Virus** - 20,462개

- **White** - 19,766개

**총 40,228개의 데이터로 구성**

→ train : 26,820 / validation : 6704 / test : 6704  
4:1:1(6 fold)

# Dataset

## Virus

- VirusShare 사이트에서 제공

**VirusShare.com** - Because Sharing is Caring


[Home](#) - [About](#) - [Hashes](#) - [Research](#) - [Support the Project](#)

Please [login](#) to search and download.

System currently contains 29,548,065 samples.

Please note that this site is constantly under construction and might be broken.

Latest sample added to the system:

	MD5	56f827366e45d22686ec09e27e7f1f3a
	SHA1	56ca4c57bd523f35159e565ba0c312fe240d698f
	SHA256	8008b2964560e2b7cc971ce08156e32a68fd6f5ba9905664cbfa02be015deb89
SSDeep	786432:ww4knqVqQHvc2tbIBfWuz5hFrblrqVBc+ClBcxD9icWx/y7fE996l2DxgioKtZUm:wwZnbQ02tweO5hFrblu7k9/wl2DjMXuD	
Size	49,710,703 bytes	
File Type	Java archive data (JAR)	
Detections	DrWeb = Adware.Leadbolt.12.origin NANO-Antivirus = Riskware.Android.Leadbolt.dkzuxh	
ExIF Data	<div>File Size : 47 MB</div> <div>File Access Date/Time : 2017:11:12 08:39:09-05:00</div> <div>File Inode Change Date/Time : 2017:11:12 06:31:08-05:00</div> <div>File Type : ZIP</div> <div>MIME Type : application/zip</div> <div>Zip Required Version : 20</div> <div>Zip Bit Flag : 0x0808</div> <div>Zip Compression : Deflated</div> <div>Zip Modify Date : 2015:04:15 14:41:19</div> <div>Zip CRC : 0x32517367</div> <div>Zip Compressed Size : 400</div> <div>Zip Uncompressed Size : 4760</div> <div>Zip File Name : assets/bin/Data/000000000000000f00000000000000</div>	
<a href="#">VirusTotal Report</a> submitted 2015-04-25 18:47:53 UTC		
VirusShare info last updated 2017-11-12 13:39:21 UTC		

# Dataset

## Virus

- 최신 파일에 대해서 torrent 피어 제공  
→ 최신 순으로 데이터 다운로드

File	Size	Added
<a href="#">VirusShare_00000.zip</a>	13.56 GB	2012-06-15 00:39:38
<a href="#">VirusShare_00001.zip</a>	85.33 GB	2012-06-15 22:12:42
<a href="#">VirusShare_00002.zip</a>	46.62 GB	2012-06-16 09:01:01
<a href="#">VirusShare_00003.zip</a>	27.57 GB	2012-06-16 09:19:46
<a href="#">VirusShare_00004.zip</a>	20.54 GB	2012-06-16 20:31:06
<a href="#">VirusShare_00005.zip</a>	22.29 GB	2012-06-17 12:48:06
<a href="#">VirusShare_00006.zip</a>	32.85 GB	2012-06-26 00:16:55
<a href="#">VirusShare_00007.zip</a>	32.85 GB	2012-06-26 00:16:55
<a href="#">VirusShare_00008.zip</a>	32.85 GB	2012-06-26 00:16:55
<a href="#">VirusShare_00009.zip</a>	32.85 GB	2012-06-26 00:16:55
<a href="#">VirusShare_00010.zip</a>	32.85 GB	2012-06-26 00:16:55
<a href="#">VirusShare_00011.zip</a>	32.85 GB	2012-06-26 00:16:55
<a href="#">VirusShare_00012.zip</a>	32.85 GB	2012-06-26 00:16:55
<a href="#">VirusShare_00013.zip</a>	32.85 GB	2012-06-26 00:16:55
<a href="#">VirusShare_00014.zip</a>	32.85 GB	2012-06-26 00:16:55
<a href="#">VirusShare_00015.zip</a>	32.85 GB	2012-06-26 00:16:55
<a href="#">VirusShare_00016.zip</a>	32.85 GB	2012-06-26 00:16:55
<a href="#">VirusShare_00017.zip</a>	31.37 GB	2012-10-26 01:08:06
<a href="#">VirusShare_00018.zip</a>	27.64 GB	2012-10-29 22:45:17
<a href="#">VirusShare_00019.zip</a>	26.10 GB	2012-10-31 14:17:58
<a href="#">VirusShare_00020.zip</a>	42.75 GB	2012-11-05 11:45:31
<a href="#">VirusShare_00021.zip</a>	59.79 GB	2012-11-20 02:07:18
<a href="#">VirusShare_00022.zip</a>	61.84 GB	2012-11-27 10:38:50
<a href="#">VirusShare_00023.zip</a>	80.84 GB	2012-11-30 23:21:49
<a href="#">VirusShare_00024.zip</a>	49.93 GB	2012-12-06 10:26:51
<a href="#">VirusShare_00025.zip</a>	46.11 GB	2012-12-19 18:21:01
<a href="#">VirusShare_00026.zip</a>	57.95 GB	2012-12-22 02:02:51
<a href="#">VirusShare_00027.zip</a>	40.43 GB	2013-01-04 17:36:04
<a href="#">VirusShare_00028.zip</a>	50.39 GB	2013-01-05 17:31:50

<a href="#">VirusShare_00267.zip</a>	17.42 GB	2016-09-01 23:34:58
<a href="#">VirusShare_00268.zip</a>	31.08 GB	2016-09-25 21:01:06
<a href="#">VirusShare_00269.zip</a>	27.83 GB	2016-10-16 12:10:01
<a href="#">VirusShare_00270.zip</a>	18.11 GB	2016-11-01 22:52:01
<a href="#">VirusShare_00271.zip</a>	12.23 GB	2016-11-20 12:55:42
<a href="#">VirusShare_00272.zip</a>	15.90 GB	2016-12-03 13:52:11
<a href="#">VirusShare_00273.zip</a>	5.96 GB	2016-12-09 22:48:00
<a href="#">VirusShare_00274.zip</a>	12.56 GB	2017-01-05 21:08:33
<a href="#">VirusShare_00275.zip</a>	12.21 GB	2017-01-05 21:16:31
<a href="#">VirusShare_00276.zip</a>	12.75 GB	2017-02-24 22:25:17
<a href="#">VirusShare_00277.zip</a>	14.37 GB	2017-02-24 22:33:46
<a href="#">VirusShare_00278.zip</a>	6.67 GB	2017-02-24 22:38:05
<a href="#">VirusShare_00279.zip</a>	13.88 GB	2017-02-24 22:51:21
<a href="#">VirusShare_00280.zip</a>	21.94 GB	2017-02-24 23:01:39
<a href="#">VirusShare_00281.zip</a>	18.71 GB	2017-03-05 14:19:43
<a href="#">VirusShare_00282.zip</a>	11.64 GB	2017-03-15 22:30:23
<a href="#">VirusShare_00283.zip</a>	16.43 GB	2017-03-15 22:50:41
<a href="#">VirusShare_00284.zip</a>	29.93 GB	2017-04-04 22:21:57
<a href="#">VirusShare_00285.zip</a>	17.65 GB	2017-04-05 23:32:00
<a href="#">VirusShare_00286.zip</a>	14.55 GB	2017-04-09 23:26:12
<a href="#">VirusShare_00287.zip</a>	14.15 GB	2017-04-26 20:57:06
<a href="#">VirusShare_00288.zip</a>	22.77 GB	2017-04-26 21:15:39
<a href="#">VirusShare_00289.zip</a>	15.12 GB	2017-05-07 13:58:50
<a href="#">VirusShare_00290.zip</a>	17.72 GB	2017-05-15 00:00:40
<a href="#">VirusShare_00291.zip</a>	25.46 GB	2017-06-01 23:10:09
<a href="#">VirusShare_00292.zip</a>	24.61 GB	2017-06-13 21:43:22
<a href="#">VirusShare_00293.zip</a>	20.41 GB	2017-07-05 00:33:53
<a href="#">VirusShare_00294.zip</a>	20.77 GB	2017-07-08 18:27:26
<a href="#">VirusShare_00295.zip</a>	32.88 GB	2017-07-24 20:26:36
<a href="#">VirusShare_00296.zip</a>	35.95 GB	2017-08-26 21:41:15

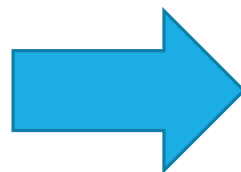


# Dataset

---

## Virus

VirusShare\_00293.zip (다운로드)  
VirusShare\_00294.zip (다운로드)  
VirusShare\_00295.zip (다운로드)  
VirusShare\_00296.zip (다운로드)



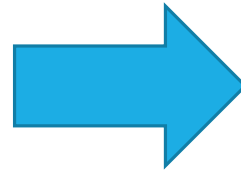
각각 65,536 개의  
malware 제공

# Dataset

## Virus

데이터 개수 : 196,608

VirusShare\_00293.zip  
VirusShare\_00294.zip  
VirusShare\_00295.zip  
VirusShare\_00296.zip



application/CDFV2-unknown  
application/gzip  
application/java-archive  
application/msword  
**application/octet-stream**  
application/pdf  
application/vnd.ms-cab-compressed  
application/vnd.ms-excel  
application/vnd.ms-office  
application/vnd.ms-powerpoint  
application/vnd.openxmlformats-officedocument...  
application/x-7z-compressed  
**application/x-dosexec**  
**application/x-elc**  
**application/x-executable**  
application/xml  
**application/x-msi**  
application/x-rar

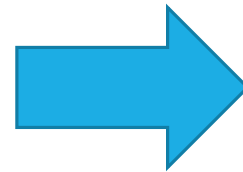
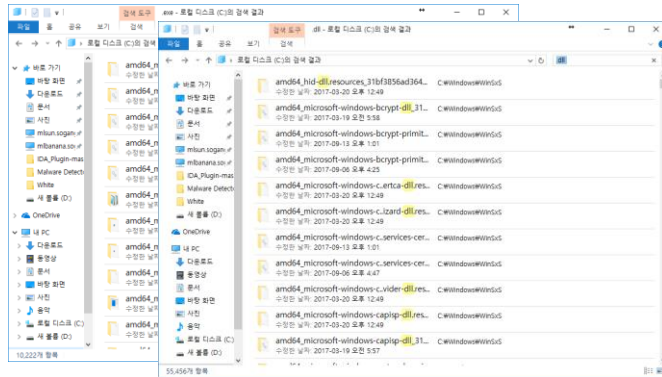
**application/x-setupscript**  
**application/x-sharedlib**  
**application/x-shockwave-flash**  
application/zip  
**application/zlib**  
image/gif  
image/jpeg  
image/png  
image/x-icon  
message/rfc822  
text/html  
**text/plain**  
**text/rtf**  
**text/rtf,application/zlib**  
**text/troff**  
**text/x-asm**  
**text/x-pascal**  
text/x-php

196,608개 → 응용프로그램 및 스크립트 유형  
**20,462**개만 사용

# Dataset

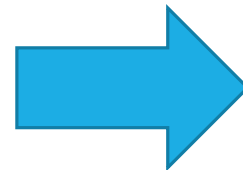
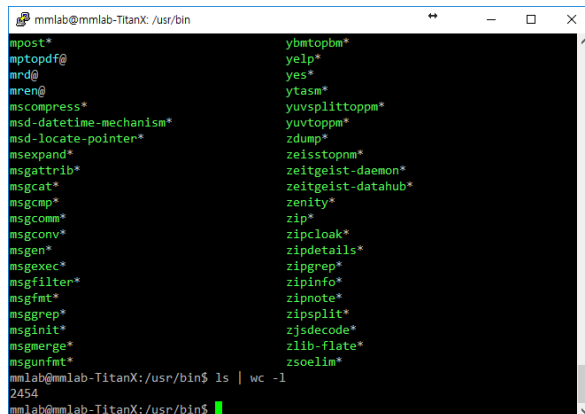
## White

Windows



\*.exe, \*.dll 파일만 추출  
→ 17,963개

Linux



Link를 제외한 실행 파일만 사용  
→ 약 1,803개

19,766개

# 시스템 구성도

---

- HW

- CPU: Intel® Core™ i7-7700 CPU @ 4.20GHz
- Main Memory: 32GB
- GPU: GeForce GTX 1080 Ti (11GB GDDR5)

- OS

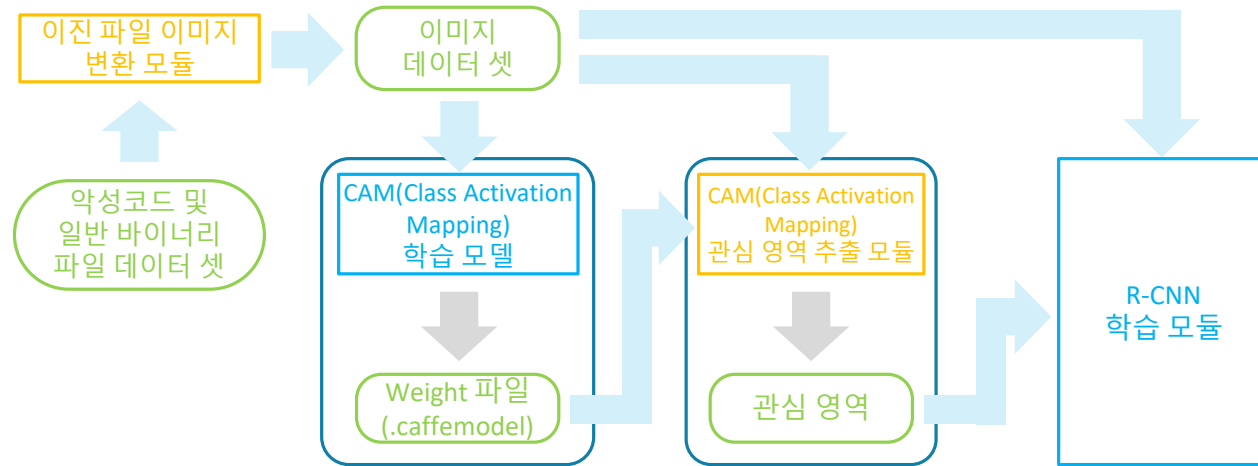
- Ubuntu 16.04.5 LTS

- System SW

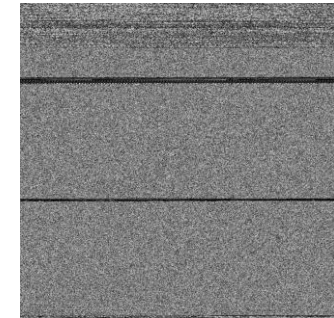
- Python: 2.7.13
- Caffe: 1.0.0
- Pycaffe: 1.0.0
- OpenCV: 2.4.9

# 시스템 구성도

## ◦ 학습 시스템

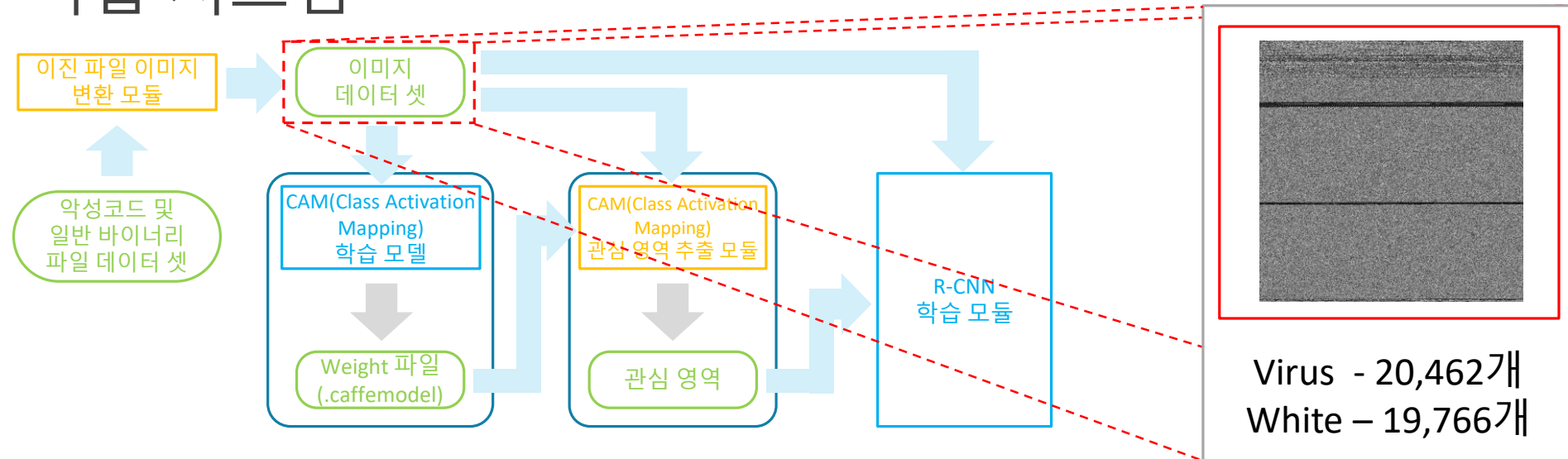


## ○ 학습 시스템

[illegible]

# 시스템 구성도

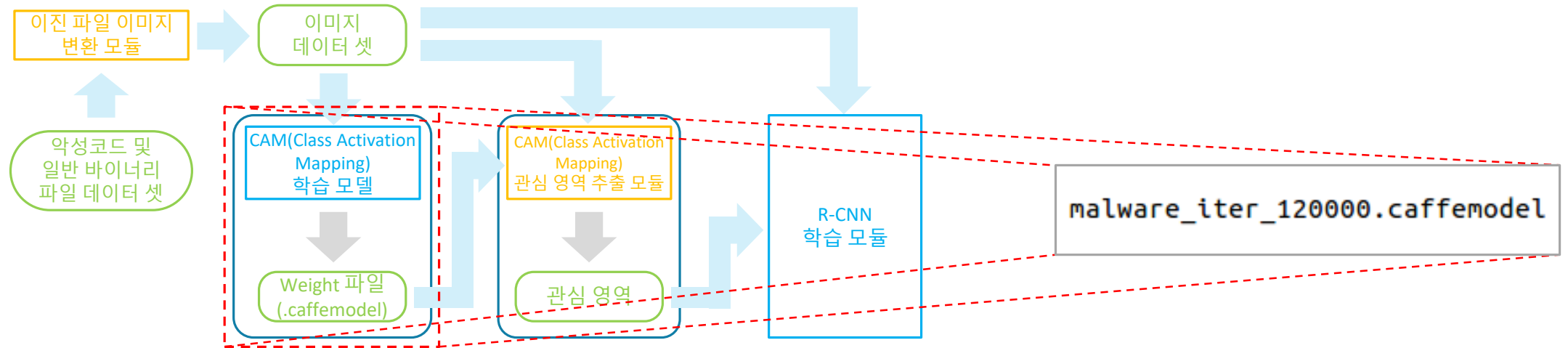
## ◦ 학습 시스템





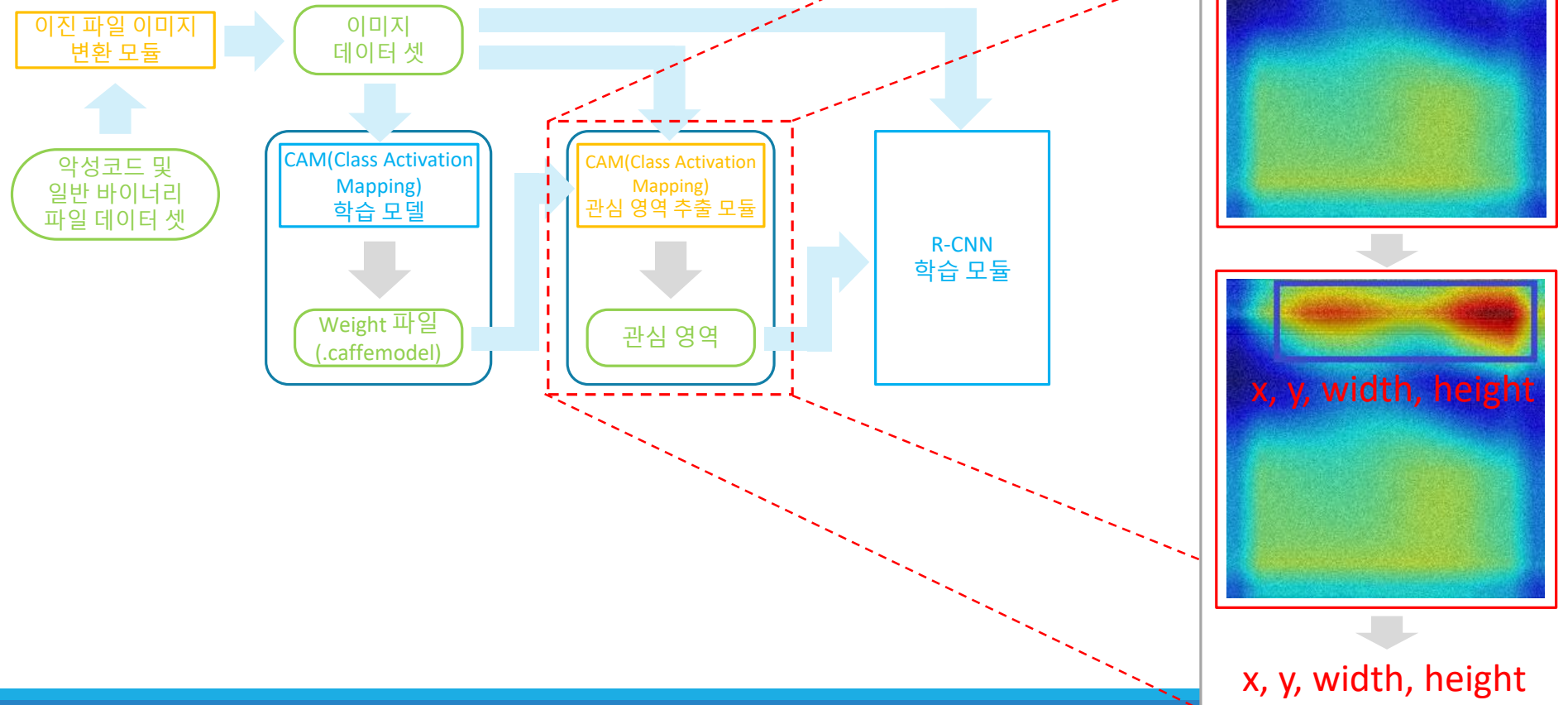
# 시스템 구성도

## ◦ 학습 시스템



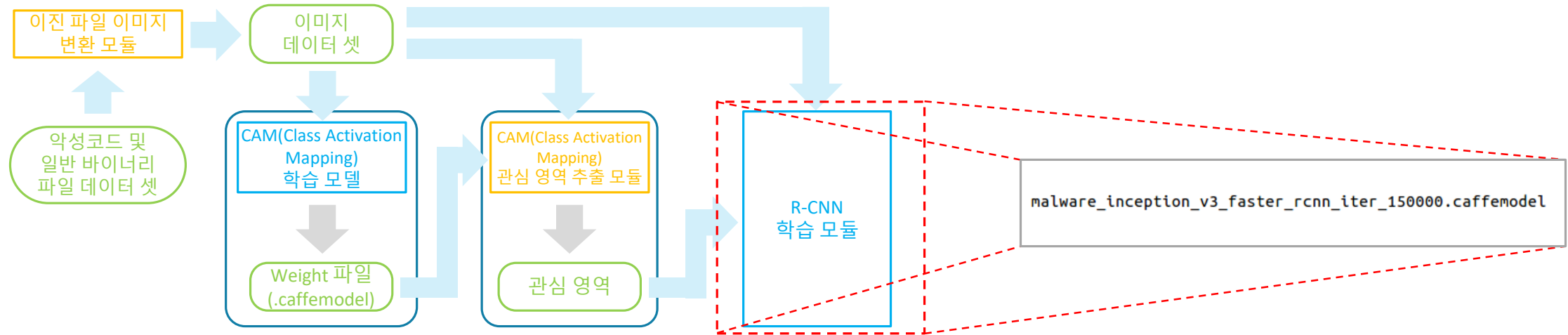
# 시스템 구성도

## ◦ 학습 시스템



# 시스템 구성도

## ◦ 학습 시스템



# 시스템 구성도

## 테스트 시스템

이진 파일 이미지  
변환 모듈

이미지  
데이터 셋

R-CNN  
테스트 모듈

새로운 파일  
(악성코드 및  
일반 파일)

virus

white

Malware  
판단 영역

malware\_inception\_v3\_faster\_rcnn\_iter\_150000.caffemodel

예제

malware

malware

# 성능

---

- Class Activation Mapping

	VGG 16	VGG 19	Inception v3	Inception v4
Recall	98.21	99.32	99.24	99.38
Precision	50.44	49.13	98.73	96.19

## 학습 옵션

test\_iter: 1000  
test\_interval: 3000  
base\_lr: 0.001  
lr\_policy: "step"  
gamma: 0.1

stepsize: 40000  
max\_iter: 160000  
momentum: 0.9  
weight\_decay: 0.0005  
snapshot: 10000

# 성능

- Class Activation Mapping

	VGG 16	VGG 19	Inception v3	Inception v4
Recall	98.21	99.32	99.24	99.38
Precision	50.44	49.13	98.73	96.19

## 학습 옵션

test\_iter: 1000  
test\_interval: 3000  
base\_lr: 0.001  
lr\_policy: "step"  
gamma: 0.1

stepsize: 40000  
max\_iter: 160000  
momentum: 0.9  
weight\_decay: 0.0005  
snapshot: 10000

# 성능

---

- R-CNN

	VGG16	Inception
Recall	99.62	<b>99.27</b>
Precision	48.17	<b>51.7</b>

## 학습 옵션

base\_lr: 0.001  
lr\_policy: "step"  
gamma: 0.1  
stepsize: 150000  
display: 100  
average\_loss: 100

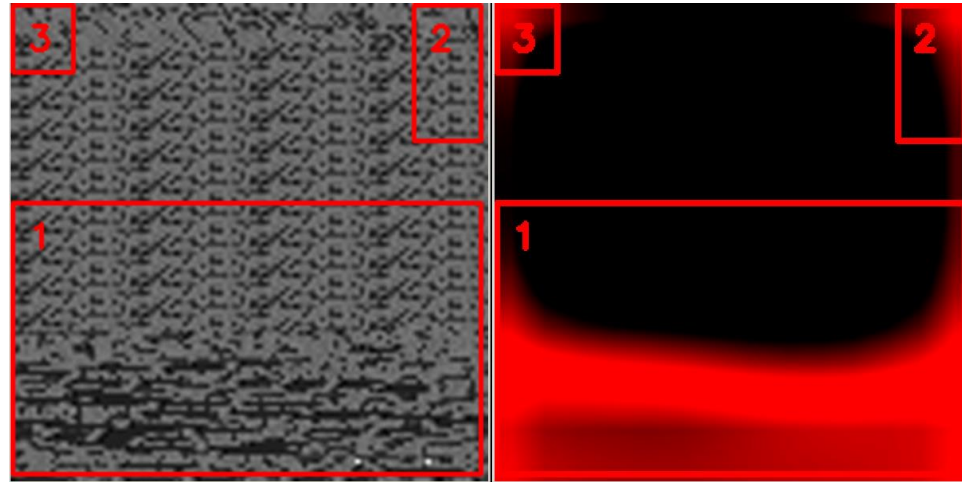
# iter\_size: 1  
momentum: 0.9  
weight\_decay: 0.0005  
snapshot: 10000  
iter\_size: 2



# 성능 : 문제점 진단

---

- Class Activation mapping



- Bonding box 생성 시 작은 부분도 상자를 생성하여 noise가 다량 포함되어 있음  
→ Filtering 필요

# 성능

- R-CNN

	VGG16	Inception
Recall	99.34	<b>98.97</b>
Precision	49.12	<b>51.04</b>

## 학습 옵션

base\_lr: 0.001  
lr\_policy: "step"  
gamma: 0.1  
stepsize: 150000  
display: 100  
average\_loss: 100

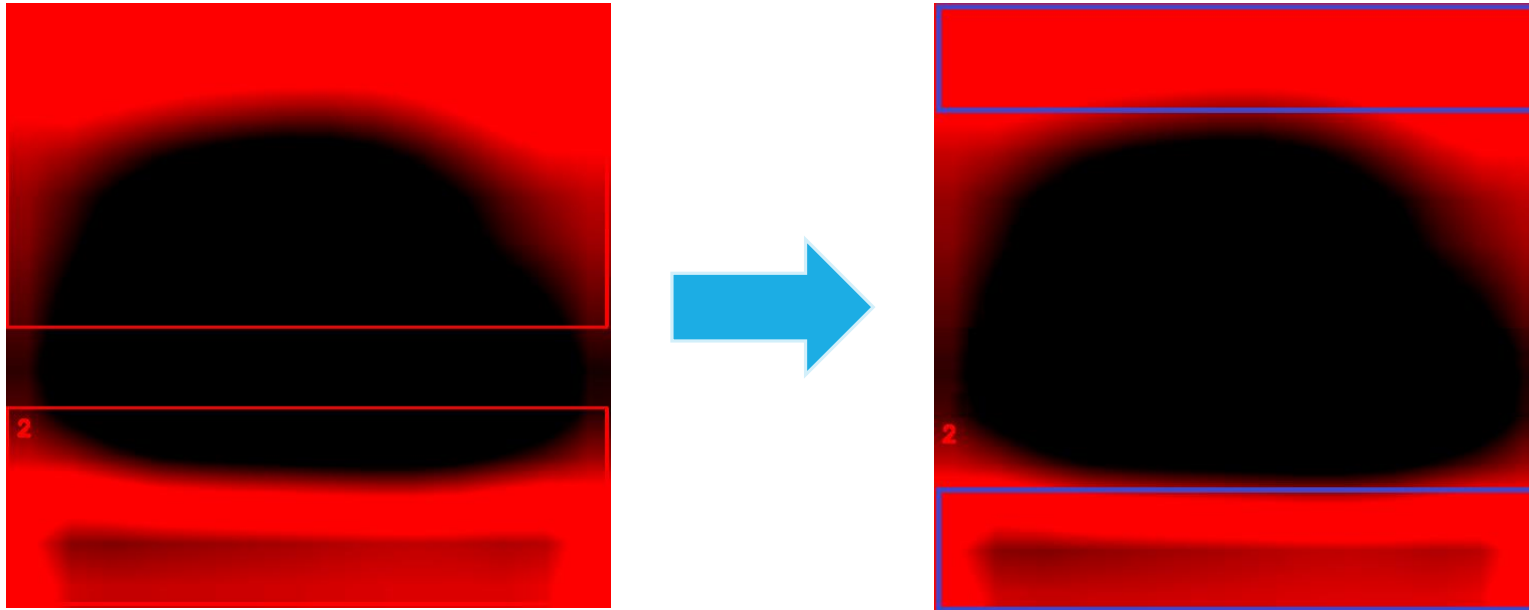
# iter\_size: 1  
momentum: 0.9  
weight\_decay: 0.0005  
snapshot: 10000  
iter\_size: 2



**성능이 개선되지 않음**

# 향후 개선 방향

---



- Bounding box에 포함되는 가장자리의 noise를 제거하여 R-CNN을 다시 학습  
→ 수작업 진행 예정