

---

박학다식

# Find Color Project

---

2019311036 신새별

2018311095 장민근

2017313764 김재연

2017314786 정동진

2015313546 김창현

# CONTENTS

---

## 01 아이디어 및 목표

1.1 연구 배경 및 목적

1.2 참고 논문 소개

## 02 AI 모델 개발 계획

2.1 Research Question

2.2 Context-aware Adaptive Network

2.3 Local & Global Hint Network

2.4 Dataset

2.5 Implementation Details

## 03 UI & UX 디자인

## 04 프로젝트 수행 계획

## 1.1 아이디어 및 목표

현재 1900년대 이미지는 흑백 이미지인 경우가 대다수이며 이러한 오래된 기록물, 콘텐츠에 대한 **colorization** 기술에 대한 needs가 증가하고 있음.

<https://m.youtube.com/watch?v=IDH7SgQtOns>



## 1.1 아이디어 및 목표

- 최근 이미지/비디오 등의 미디어 데이터에 AI 기술을 적용하여 품질을 향상시키는 기술이 발전되고 있으며 흑백 이미지에 컬러를 입히는 **Colorization 기술**을 통해 이러한 자료들을 **복원하는 것은 사회/경제적 가치 측면에서 매우 의미** 있는 일이라고 할 수 있음.
- 오래된 기록물, 콘텐츠 등을 대상으로 대한민국 역사를 담고 있는 디지털 흑백 이미지에 컬러를 입히는 AI 복원 기술 개발 및 이미지가 담고 있는 **상황 (전쟁, 해방, 시위) 및 특성에 맞는 Adaptive한 Colorization model 개발을 목표**로 함.



Colorization



## 1.2 참고 논문 소개

---

### 1. Real-Time User-Guided Image Colorization with Learned Deep Priors

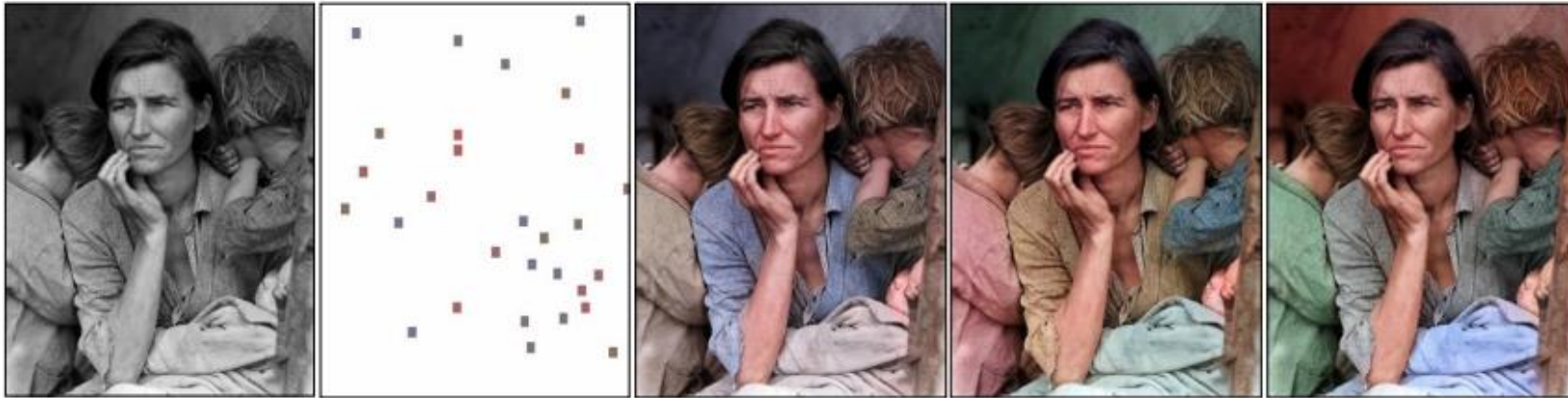
- 최근 Zhang et. al. (2016) 이나 Larsson et al. (2016) 등 다양한 저자들이 fully automatic colorization을 활용하여 cheap and easy color image를 생성했지만, artifacts들이 존재하거나 적합하지 않은 색을 colorization하는 현상이 일어남.
- 이 논문에서는 Local Hint Network, Global Hint Network를 사용하여 main Network에 적용하여 “hint”를 주어 colorization의 정확도를 높이는 방법을 제안함.



## 1.2 참고 논문 소개

### Local Hint Network

- Sparse user point와 input grayscale image를 concat한 뒤에 input으로 넣음. 각 pixel마다 color distribution을 예측하여 user에게 색을 추천함.

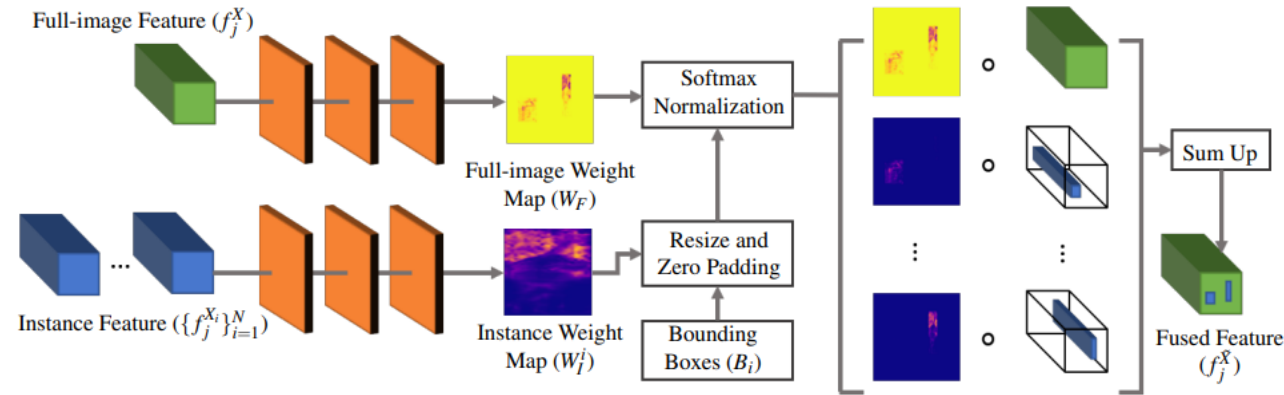


### Global Hint Network

- 사용자가 global statistics(histogram, saturation)을 제공.
- ground truth의 색 분포, saturation을 학습할 때 random으로 hint로 줌.  
(주거나 주지 않을 수 있음)

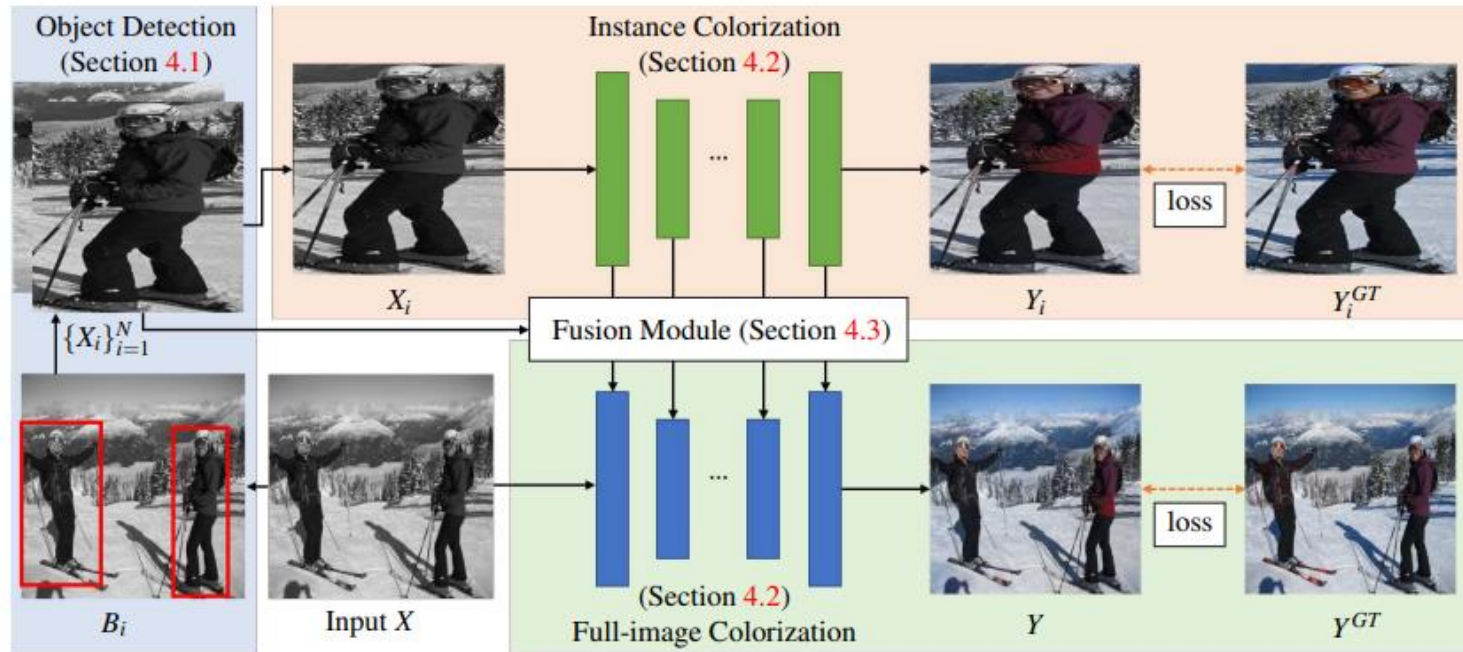
## 1.2 참고 논문 소개

### 2. Instance-aware Image Colorization



- 전통적인 colorization 방법은 color scribbles, reference image 같은 가이드를 제공하는 것처럼 사용자의 개입에 의존함.
- 현존하는 방법인 deep neural network는 learning과 colorization을 전체 이미지에 대해 수행하기 때문에 object가 여러 개가 있을 경우 성능이 좋지 못한 문제가 있음.
- 이 논문에서는 “Clear Figure-Ground Separation”을 통해 colorization 성능을 향상시키는 방법을 제안하였음

## 1.2 참고 논문 소개



- 먼저 Off-the-shelf pre-trained object detector를 활용하여 흑백 이미지로부터 multiple object bounding boxes를 얻어낸 후 instance image들의 세트를 생성하여 detected bounding boxes를 이용해 흑백 이미지로부터 잘라낸 이미지를 resize.
- 다음으로, instance image  $X_i$  는 instance colorization network에, input grayscale image  $X$ 는 full-image colorization network에 feed한다.
- 마지막으로, fusion module을 이용하여 모든 instance features를 full-image feature와 융합



## 1.2 참고 논문 소개

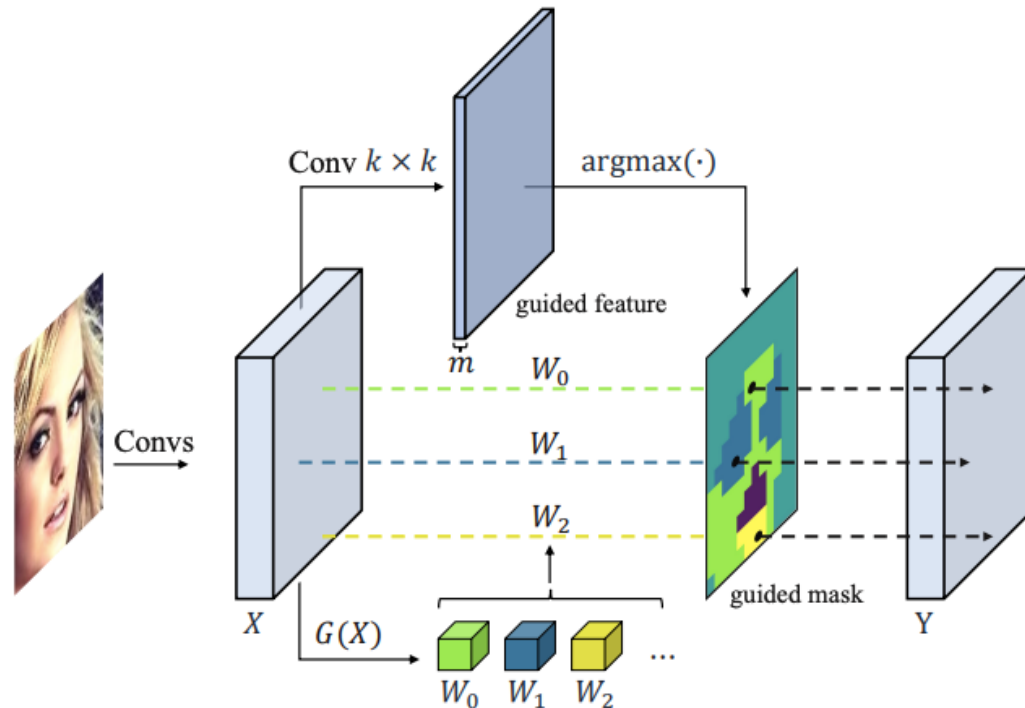
---

### 3. Dynamic Region-Aware Convolution

- 기존의 standard convolution 기법의 경우, spatial domain 간의 filter sharing을 이용하여 진행하기 때문에 효과적인 정보를 얻기 위해서는 더 많은 필터로 채널과 깊이를 증가시켜야 하고, 연산이 반복적으로 적용되어야 했음.
- 이 방법은 계산 비용이 많이 들고, 최적화의 어려움을 유발시킬 수 있음.
- Dynamic Region-Aware Convolution(DRconv)는 learnable guided mask를 활용하여 spatial dimension region에 필터를 자동으로 할당하기때문에 object를 인식하는 능력과 의미적인 정보를 파악하는 능력이 뛰어남.

## 1.2 참고 논문 소개

- standard convolution을 이용하여 guided feature를 생성  
guided feature에 따라 spatial dimension을 몇몇 영역으로 나눔
- 의미적으로 비슷한 feature들은 guided mask내 같은 영역으로 배정
- 각각의 shared region에서 filter 들을 이용하여 2D convolution 진행
- 모든 filter를 공유하는 standard convolution 방식과는 다르게 일부만 각각의 영역에서 공유하기 때문에 computational cost에 대한 이점과 semantic information에 대한 이점 또한 가짐



## 2.1 Research Question

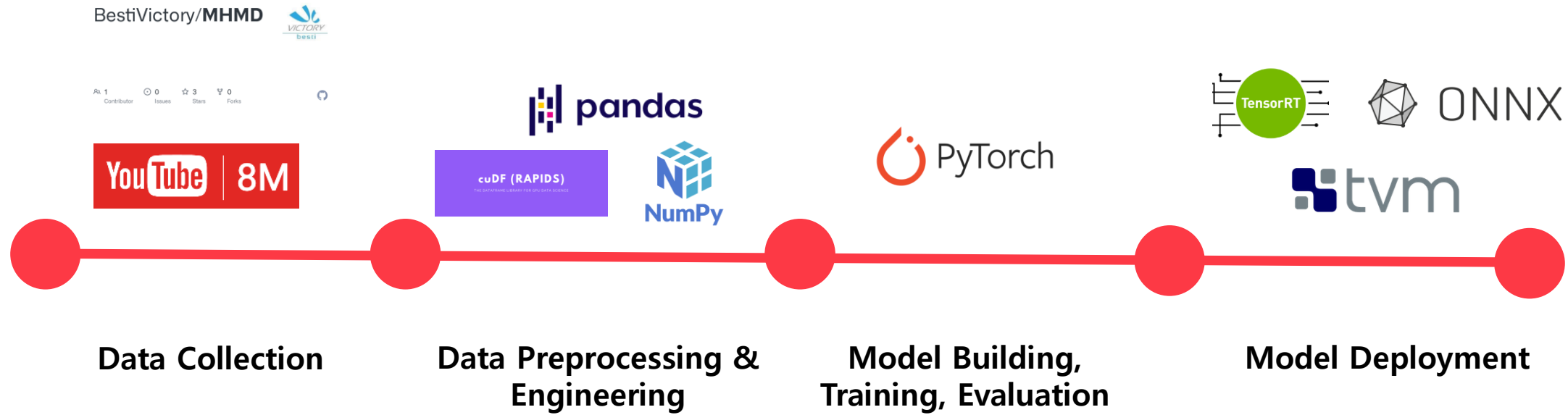
---

RQ1. 흑백데이터만 있는 경우 효과적으로 학습을 진행하는 방법은?

RQ2. 상황에 맞는 'Adaptive' 한 모델을 어떻게 만들 것인가?

RQ3. 제안하는 방법은 현업에서 쓰이기 Efficient 한가?

# 전체 프로세스 도식화



# RQ1. 흑백 데이터만 있는 경우 효과적으로 학습을 진행하는 방법은?

## MHMD (Modern Historical Movies Dataset)

- 오래된 영화와 TV 시리즈 147편에서 전처리 한 **다양한 의류 유형, 시대 및 국적**의 요구사항을 동시에 충족시키는 1,353,166개의 이미지 데이터셋
- 현존하는 데이터셋에는 **역사적인 흑백 사진**이나 **인물의 의상 색상**에 대한 **정보가 부족**하다는 문제를 해결 가능
- 오래된 기록물을 **왜곡없이 복원**하는데 적합한 데이터 셋이라고 할 수 있음

(<https://arxiv.org/pdf/2108.06515.pdf>)



	The labels of the dataset							
Dataset	Scene	Era		Nationality		Garment Type		Total
MHMD	×	Before WWII	66,900	Chinese	753,473	Military	707,771	1,353,166
				American	934,415			
		During WWII	547,318	Russian	45,291	Formal	104,763	
				German	59,015			
		After WWII	738,948	Japanese	110,562	Informal	540,632	
				English	46,641			
ImageNet [11]	✓	×		×		×		about 1,300,000
COCO-Stuff [4]	✓	×		×		×		about 164,000
Places205 [45]	✓	×		×		×		20,500



## RQ1. 흑백 데이터만 있는 경우 효과적으로 학습을 진행하는 방법은?

### YouTube-8M Dataset

- YouTube에는 다양한 **상황, 시대, 국적** 등의 풍부한 정보가 담겨있음
- **시대**에 맞는 Colorization, **상황**에 맞는 Colorization, **인물의 국적**에 따른 Colorization 등 모델을 효과적으로 학습 가능.
- 과거의 영상을 **현대적으로 재해석**하려는 경우에도 적합할 것으로 보임
- 크롤링을 통해 부족한 데이터는 보충 가능

#### YouTube-8M Dataset

YouTube-8M is a large-scale labeled video dataset that consists of millions of YouTube video IDs, with high-quality machine-generated annotations from a diverse vocabulary of 3,800+ visual entities. It comes with precomputed audio-visual features from billions of frames and audio segments, designed to fit on a single hard disk. This makes it possible to train a strong baseline model on this dataset in less than a day on a single GPU! At the same time, the dataset's scale and diversity can enable deep exploration of complex audio-visual models that can take weeks to train even in a distributed fashion.

Our goal is to accelerate research on large-scale video understanding, representation learning, noisy data modeling, transfer learning, and domain adaptation approaches for video. More details about the dataset and initial experiments can be found in our [technical report](#) and in previous workshop pages ([2018](#), [2017](#)). Some statistics from the latest version of the dataset are included below.

**6.1 Million**  
Video IDs

**350,000**  
Hours of Video

**2.6 Billion**  
Audio/Visual Features

**3862**  
Classes

**3.0**  
Avg. Labels / Video

The videos are sampled uniformly to preserve the diverse distribution of popular content on YouTube, subject to a few constraints selected to ensure dataset quality and stability:

- Each video must be public and have at least 1000 views
- Each video must be between 120 and 500 seconds long
- Each video must be associated with at least one entity from our target vocabulary
- Adult & sensitive content is removed (as determined by automated classifiers)

## RQ2. 상황에 맞는 'Adaptive' 한 모델을 어떻게 만들 것인가?

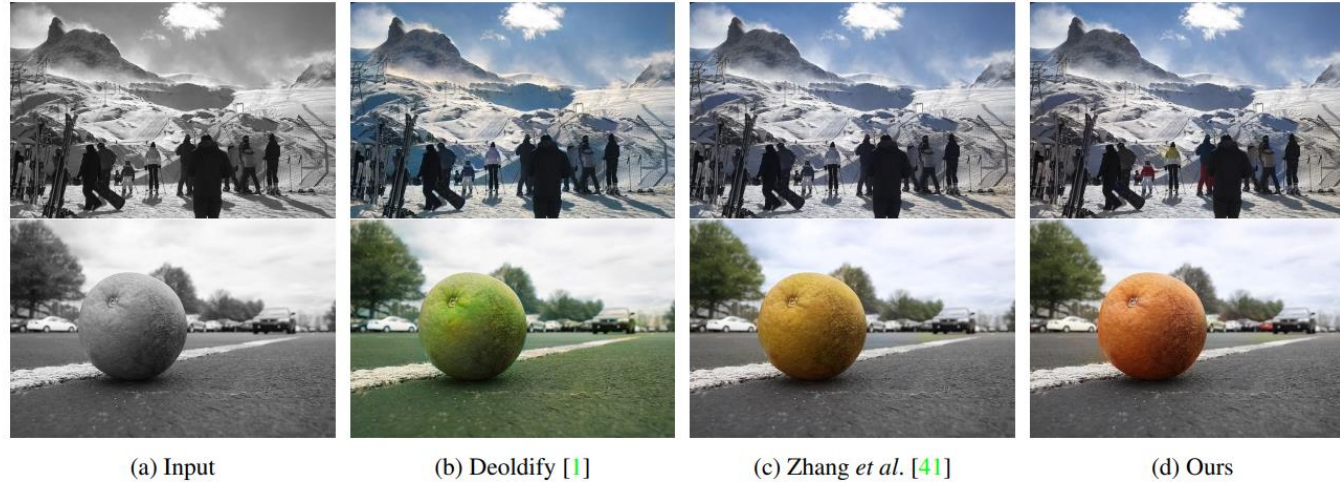


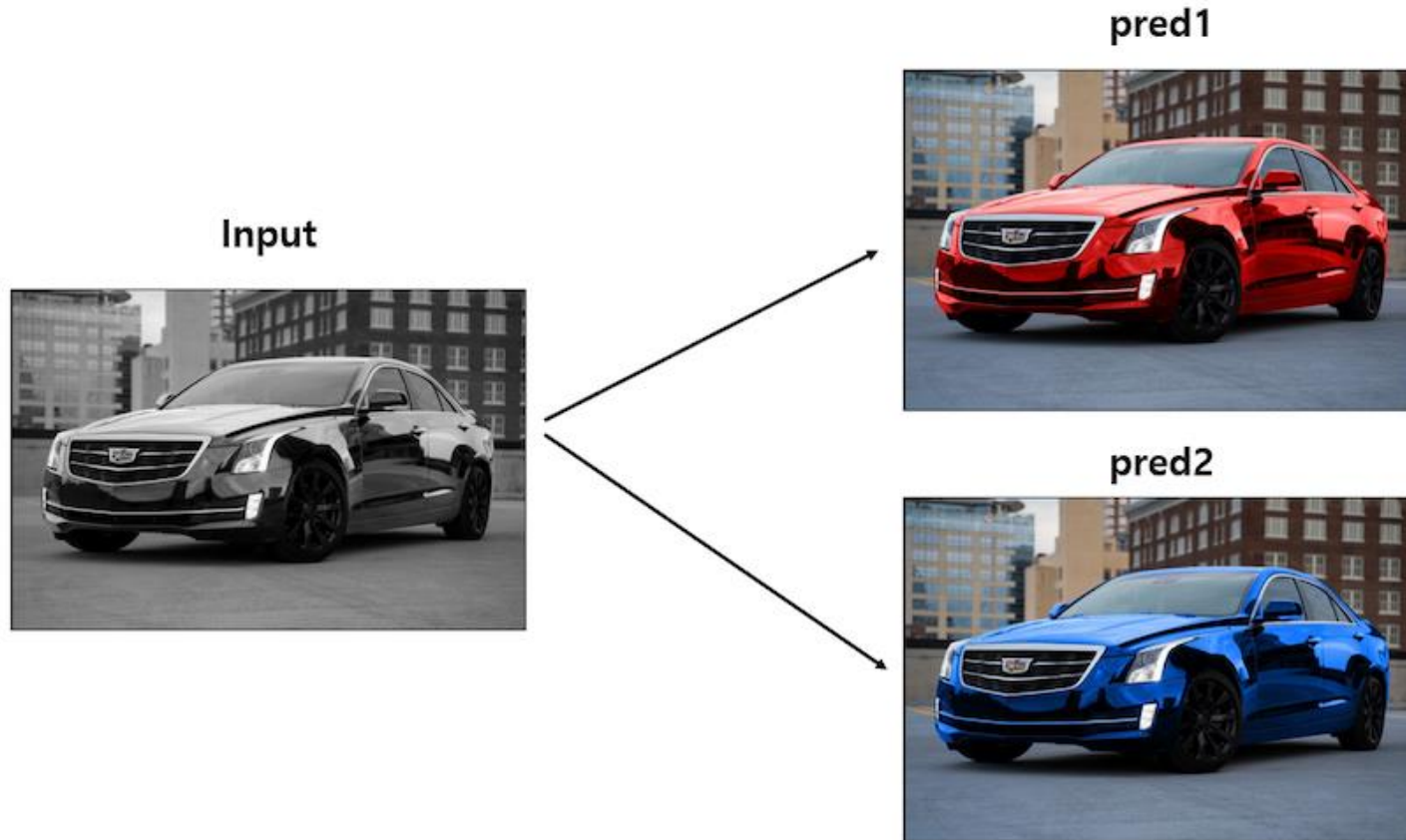
Figure 2. **Limitations of existing methods.** Existing learning-based methods fail to predict plausible colors for multiple object instances such as skiers (top) and vehicles (bottom). The result of Deoldify [1](bottom) also suffers the context confusion (biasing to green color) due to the lack of clear figure-ground separation.

- 'Deoldify' 는 context confusion을 하기에 bias된 colorization을 하게됨
- Instance-aware Image colorization [CVPR'20] 에서는 instance feature와 full-image feature를 고려하여 context에 맞는 colorization을 함

context & instance adaptive model 필요

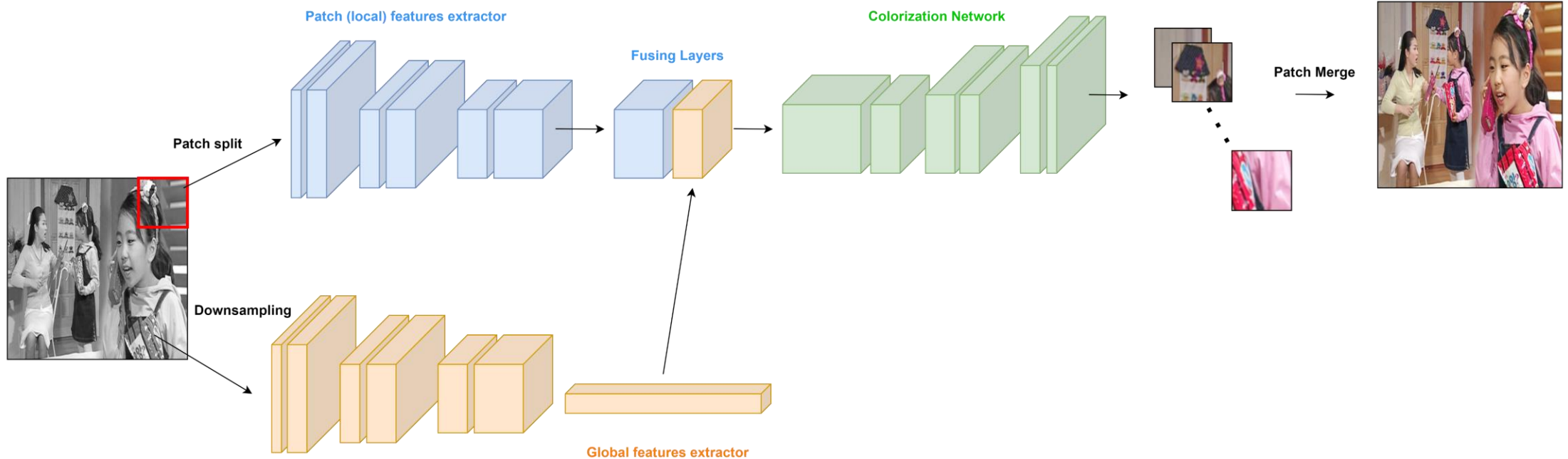
## RQ2. Challenges & Solution - ill-posed problem

User interactive colorization 필요



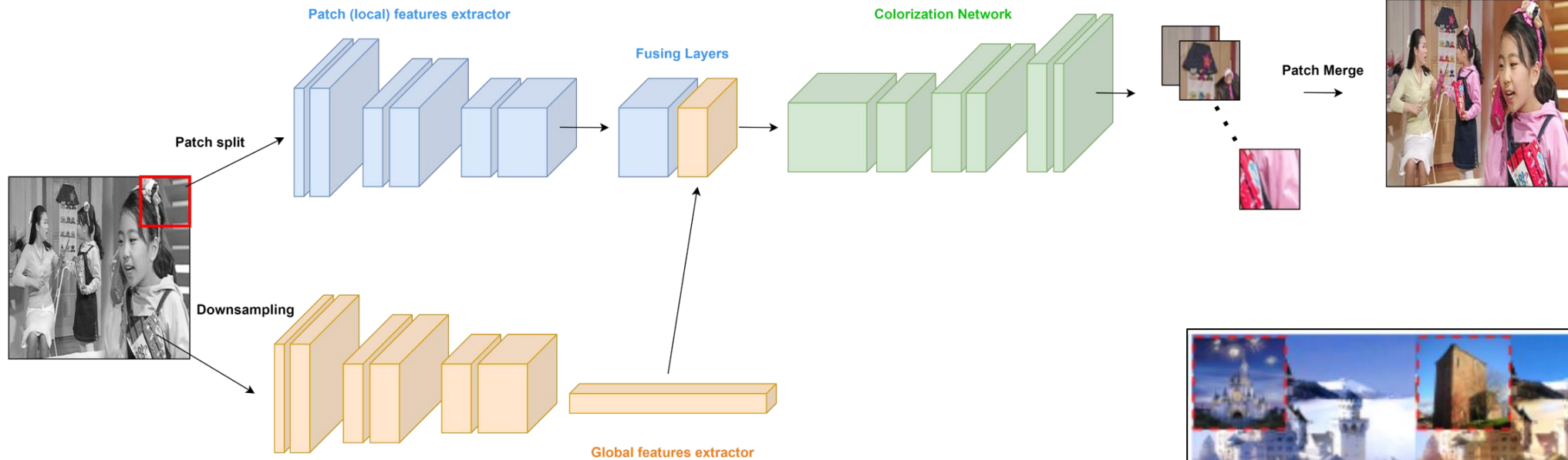
## RQ2. 상황에 맞는 'Adaptive' 한 모델을 어떻게 만들 것인가?

- 모델의 구조는 **기존 연구들의 module**과 **방법론**을 **쉽게 적용하기 위해** Unet과 같은 모델을 사용 (Instance aware colorization, Colorization Transformer ...)
- 큰 영상의 이미지를 **patch 단위로 처리**할때 생기는 문제는 **global features extractor**를 사용할 예정

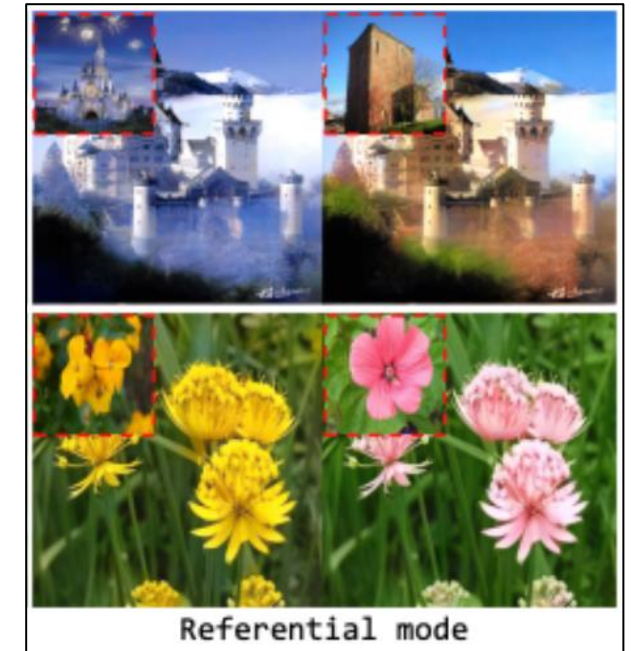




## RQ2. 상황에 맞는 'Adaptive' 한 모델을 어떻게 만들 것인가?

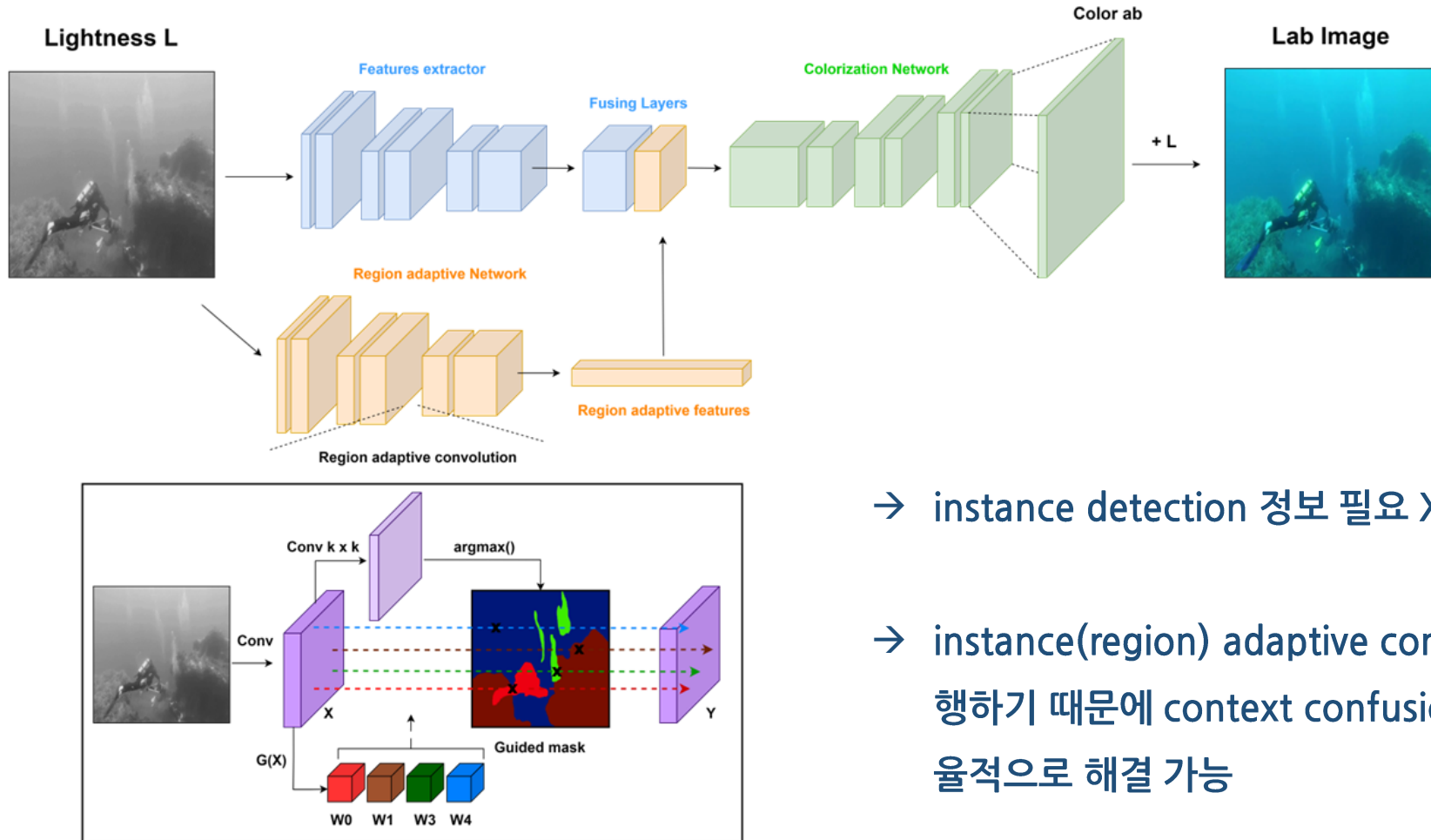


- SCSNet [AAAI'22] 를 참고하여 **referential mode** colorization 방법 도입 예정
- Source와 reference 이미지 간의 정보를 보다 **효과적으로 aggregate하는 모듈** 사용
- 기존의 colorization 에 비해 bias된 결과를 줄일 수 있음





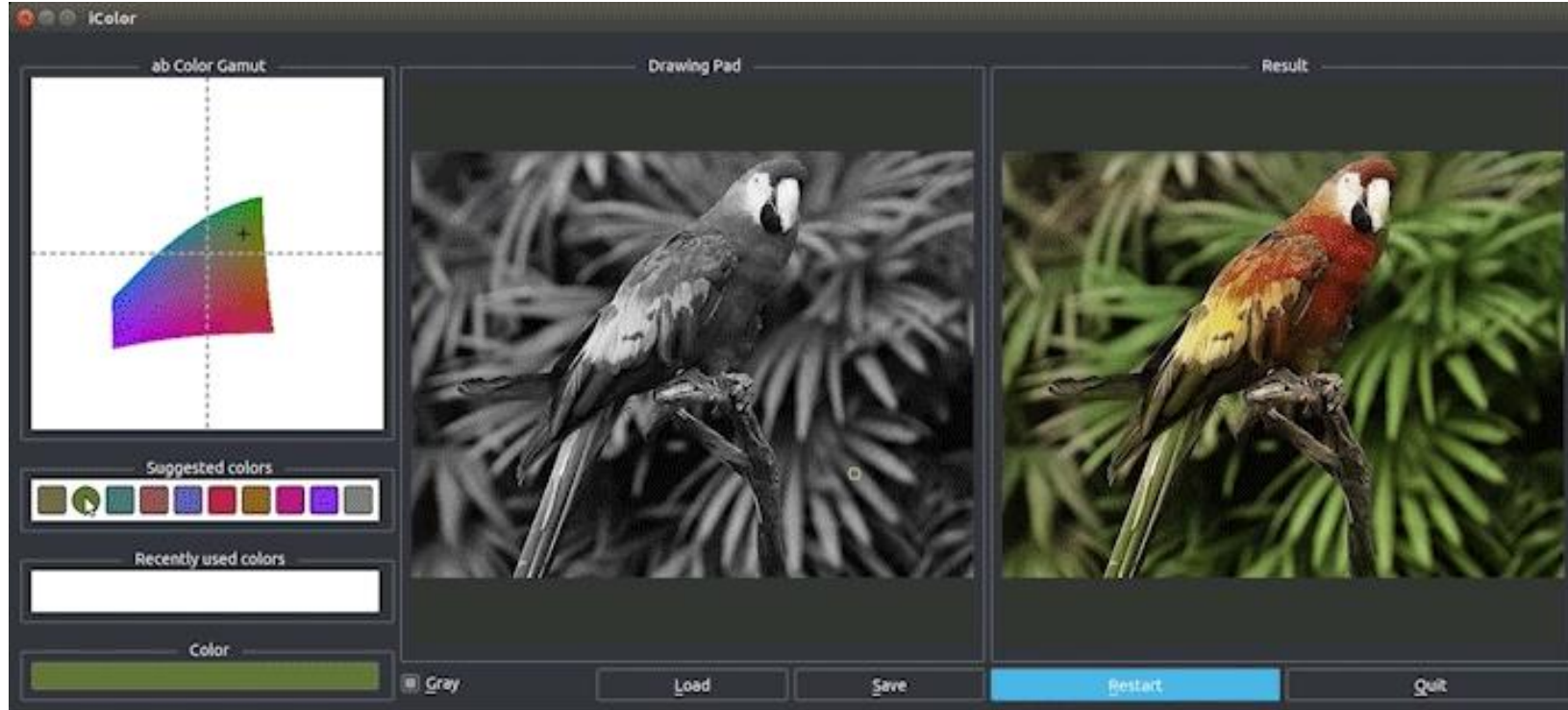
## 2.2 Our Method - Context-aware Adaptive network



→ instance detection 정보 필요 X

→ instance(region) adaptive convolution을 수행하기 때문에 context confusion 문제를 더 효율적으로 해결 가능

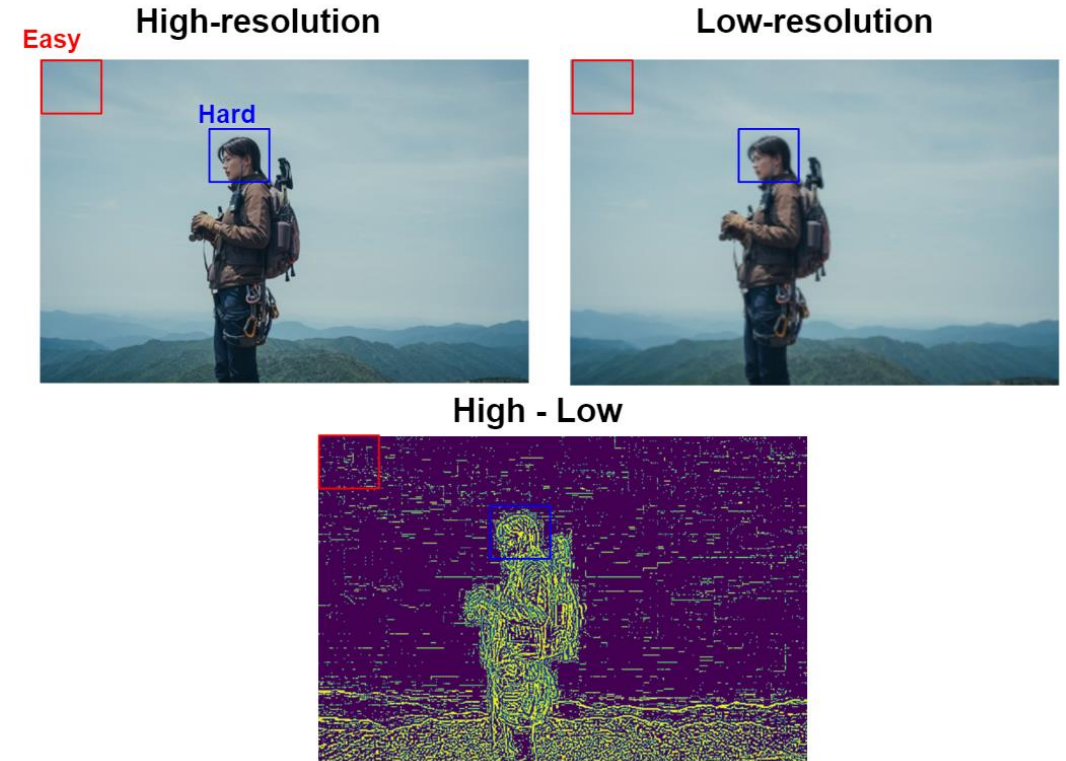
## 2.3 Our Method - Local & Global Hint Network



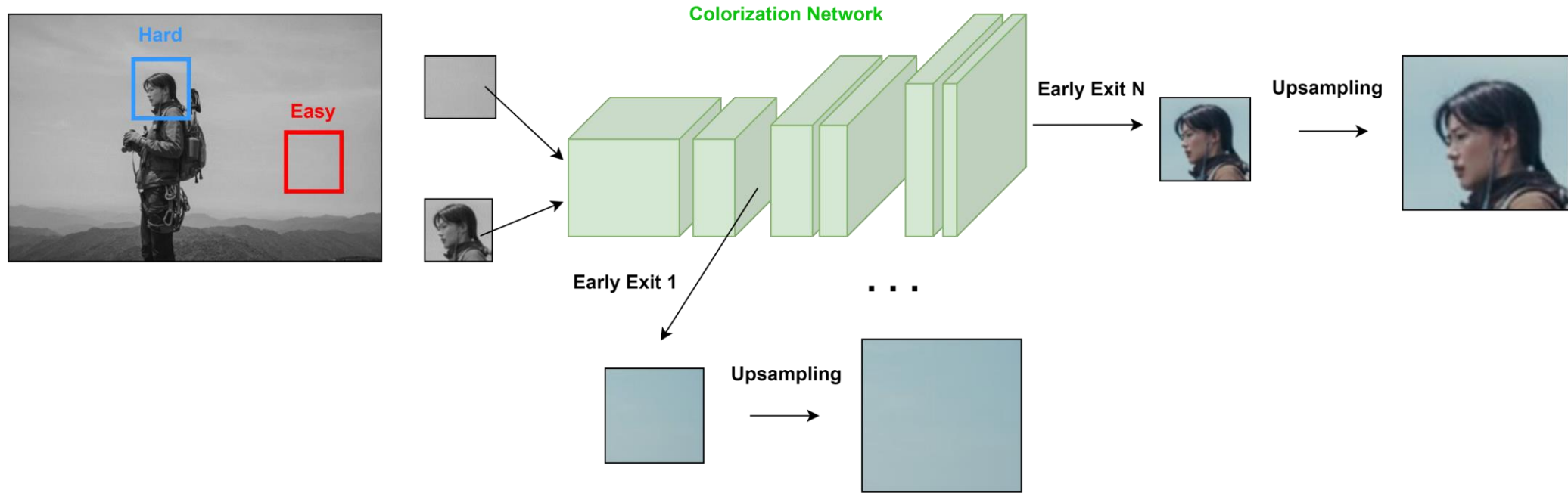
- ill-posed problem를 겪을 수밖에 없는 image colorization task에서 사용자가 개입하여 색상 변경
  - 상황에 더 적절하고 정확한 colorization 수행 가능

### RQ3. 제안하는 방법은 현업에서 쓰이기 Efficient 한가?

- Patch 단위로 연산 진행할 경우 **region에 따라 연산 효율화** 가능
- Difficulty Classification Module (DCM)을 통해 **patch-wise 난이도 분류**
- Colorization 또한, DCM 학습을 통해 연산 효율화 가능할 것으로 보임



### RQ3. 제안하는 방법은 현업에서 쓰이기 Efficient 한가?



- Patch단위로 큰 연산이 필요 없는 영역은 **Early Exit**을 통해 연산비용 압축
- 더 세밀한 Colorization이 필요한 경우 더 많은 연산 진행
- Patch 단위로 이미지 생성 후, 이미지 merge

## 2.4 Dataset

---

### ImageNet

- 100만장 가량의 이미지 데이터셋
- 클래스 1000개

### ActivityNet

- 2만개 가량의 동영상 클립
- 200가지 유형의 활동과 youtube에서 수집한 총 849시간의 동영상 포함

### Kinetic400

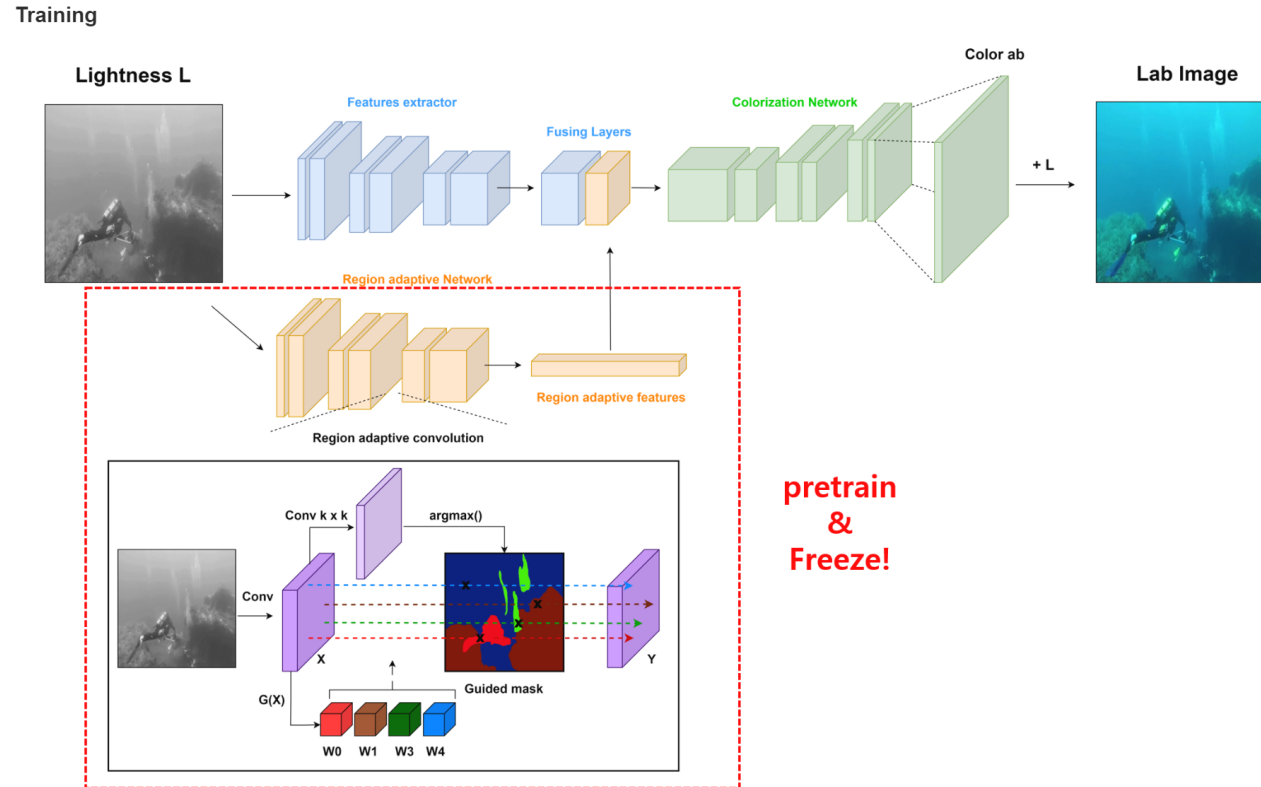
- 30만개 가량의 동영상 클립
- 클래스 400개

→ Image colorization을 효과적으로 수행하기 위해서는 **많은 데이터로 학습시킨 pretrain model이 필요**

다양한 object와 상황을 포함하는 데이터셋을 구축하기 위해 **ImageNet, Activitynet, Kinetic400 데이터를 활용할 예정**



## 2.5 Implementation details



- Context-aware adaptive network가 학습되고 나면 Local and Global Hint Network 의 방법론을 적용하여 user-interactive하게 세세한 색상을 변경할 수 있도록 추가로 학습할 예정

# Summary & Limitation

---

## < Summary >

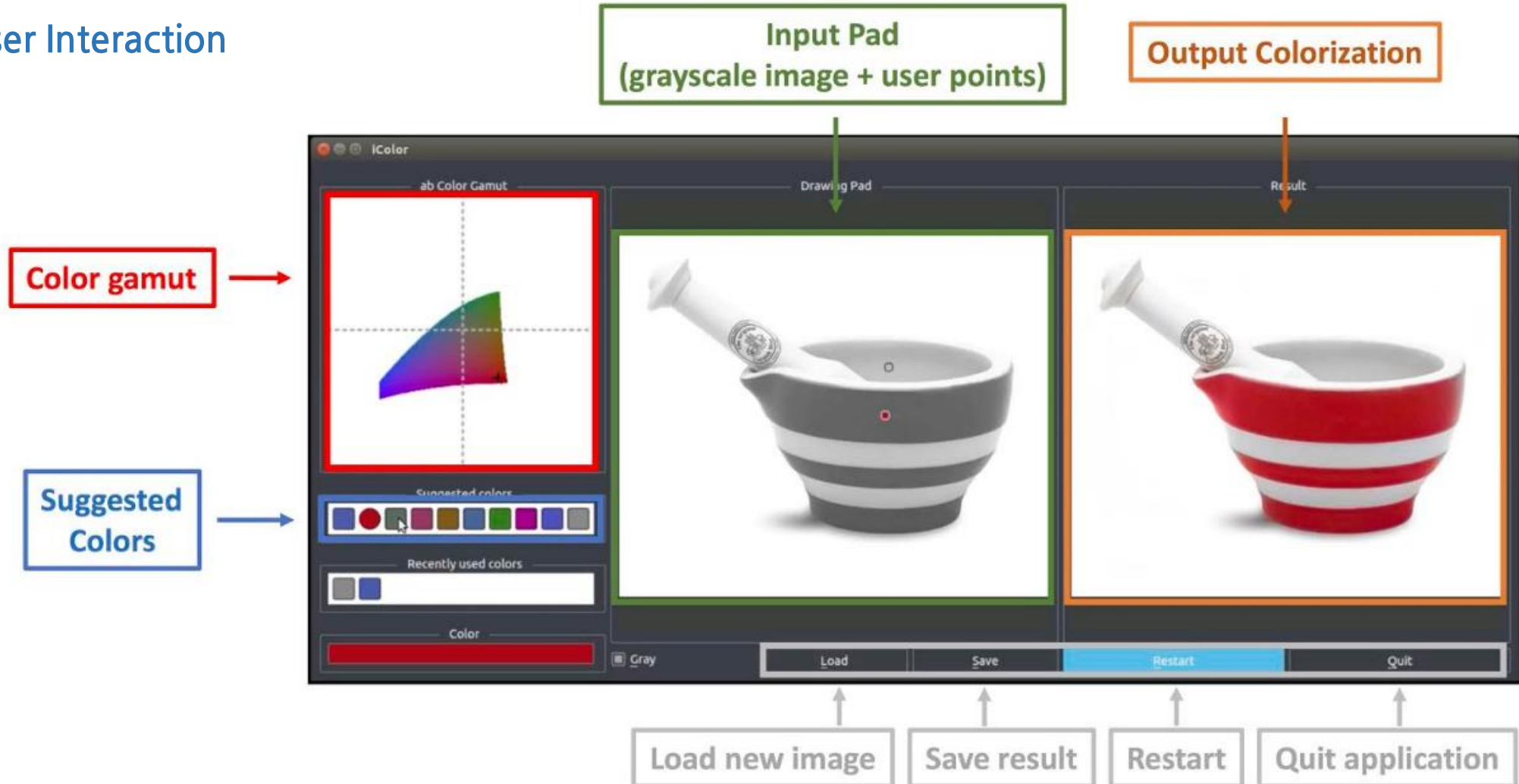
- Youtube, MHMD 데이터셋을 활용해 **흑백-컬러 pair** 데이터셋 구성
- Adaptive Model을 만들기 위해 **Global, Instance, Patch features** 를 고려
- Reference-based method 도입을 통해 복원성능 극대화
- Efficient한 모델 만들기 위해 **Adaptive Inference** 모델 제안

## < Limitation >

- 오래된 영상에서 발생하는 노이즈를 함께 해결해야 할 것으로 보임 (SR + Colorization)
- GAN-based approach

### 3. UI & UX 디자인

#### User Interaction




# find color Project


ad color Gamut



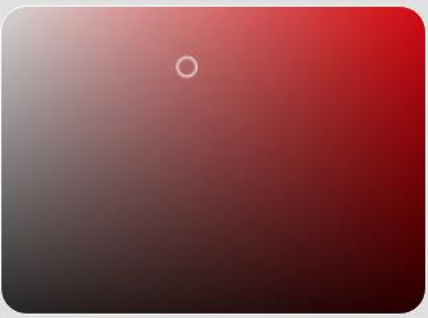
suggested colors




suggested colors



Color




Drawing Pad



Load

Save

Result Image



Restart

Quit

[illegible]



감사합니다

---