# LayoutNet, HorizonNet, HoHoNet

2021. 09. 07

AI융합학부 길다영

https://github.com/arittung/3D_Room_Reconstruction

# CONTENTS



# 1 LayoutNet

https://github.com/sunset1995/pytorch-layoutnet

# 2 HorizonNet

https://github.com/sunset1995/HorizonNet

# 3 HoHoNet

https://github.com/sunset1995/HoHoNet

# 3 HoHoNet

## HoHoNet

높이 치수가 평평한 잠재 수평 특징 표현(LHFeat)을 통해 레이아웃 구조, 조밀한 깊이 및 의미 분할을 모델링하기 위한 새로운 딥 러닝 프레임워크.

## HoHoNet은 두 가지 측면에서 발전했다

① 심층 아키텍처는 향상된 정확도로 더 빠르게 실행되도록 재설계됨.

② LHFeat의 픽셀당 조밀한 예측을 가능하게 하여 열당 출력 형태 제약을 완화하는 새로운 수평선 대 밀도 모듈을 제안함.

## 궁금한 점.

① 열당 예측과 픽셀당 예측의 차이?

② HoHoNet 실행 시, Depth estimatio와 semantic sementation의 결과가 도출되고 이 두가지 결과로 3d room reconstruction 하는 것 같다. 그렇다면, 논문의 원리 설명에서 이 두가지의 실행 구간은 어디인가..?

③ Hohonet은 Depth estimatio와 semantic sementation이 나오고 그걸 horizonnet처럼 3d room reconstruction을 하는 듯..? → 이걸 코드에서 알 수 있을까?

# 3 HoHoNet

## 내용 정리

Hohonet의 결과 : depth estimation, sematic segmentation, room layout estimation

Hohonet은 레이아웃 재구성을 위한 새로운 방법을 설계하는 것이 아님.

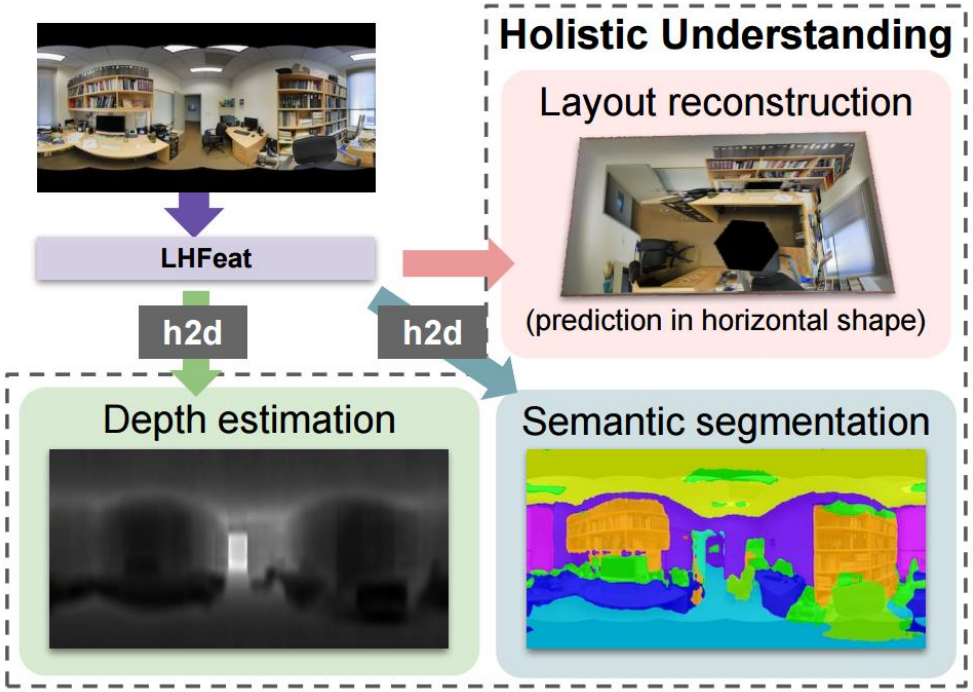
**Room layout** : Hohonet은 resnet34, horizonnet은 resnet50 사용.

**Semantic segmentation** : resnet101 사용.

**Depth estimation** : resnet50 사용

- HoHoNet은 이전의 최첨단 기술인 BiFuse[24]를 큰 폭으로 능가한다는 것을 입증한다. HoHoNet은 장면의 전체적인 구조를 잘 포착한다.
- BiFuse는 ERP와 큐브맵을 모두 모델 입력으로 사용하므로 두 개의 backbone 네트워크가 필요.
- Hohonet 단점 : HoHoNet의 깊이 경계는 BiFuse의 깊이 경계에 비해 더 흐릿하다. 열의 일부 고주파 신호는 HoHoNet에 의해 폐기된다.

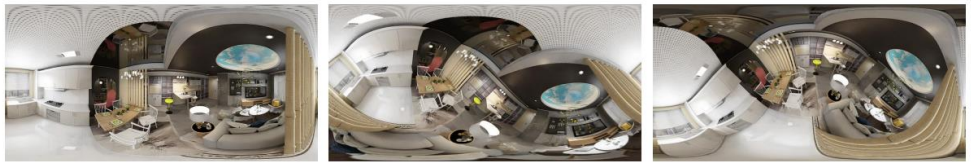
# 3 HoHoNet - 소개




① 입력 ERP 이미지는 형상 피라미드 추출을 위해 먼저 CNN backbone을 통과한 다음,

② 제안된 효율적인 높이 압축 모듈은 형상 피라미드를 높이 치수가 평평한 잠재 수평 형상 표현(LHFeat)으로 인코딩한다.

③ 마지막으로, LHFeat에서 HoHoNet 프레임워크는 최첨단 품질의 열당 및 픽셀당 양식(레이아웃의 모서리 또는 경계)을 모두 제공할 수 있다.

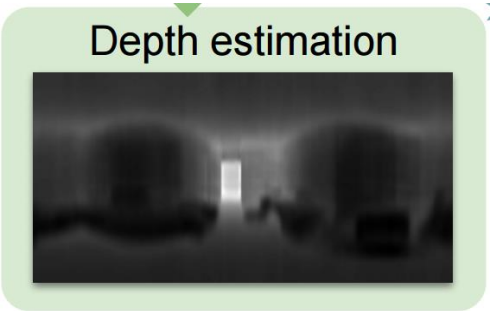


(a) Aligned 360. (b) Roll rotation. (c) Pitch rotation.

➞ (a)처럼 y축이 중력방향으로 정렬되었을 때 이미지 column 구조 정보를 더 잘 압축하여 보관할 수 있음.

# 3 HoHoNet - 소개

## 360 이미지에 대한 깊이 추정.



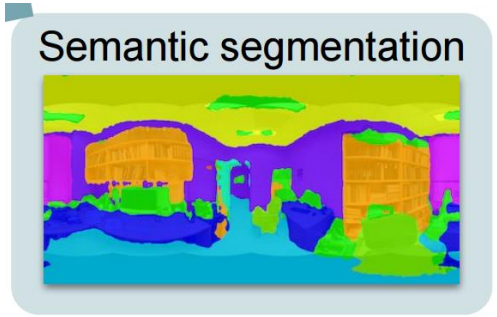

전방위 이미지의 깊이를 모델링하기 위해, OmniDepth는 ERP 왜곡을 고려하여 encoder-decoder architecture를 설계한다.

PanoPopups는 평면 인식 손실로 360 깊이를 학습하는 것이 합성 환경에 도움이 된다는 것을 보여준다.

계단식 훈련 단계를 가진 여러 백본(backbone)을 사용하는 대부분의 최신 방법과 대조적으로, HoHoNet은 하나의 백본으로만 구성되며 한 단계에서만 훈련된다.

또한, HoHoNet은 소형 LHFeat을 통해 밀도가 높은 깊이를 모델링하는 반면 이전의 기술은 기존의 밀도가 높은 특징에서 깊이를 추정한다.

## 360 이미지에 대한 의미론적 세분화.



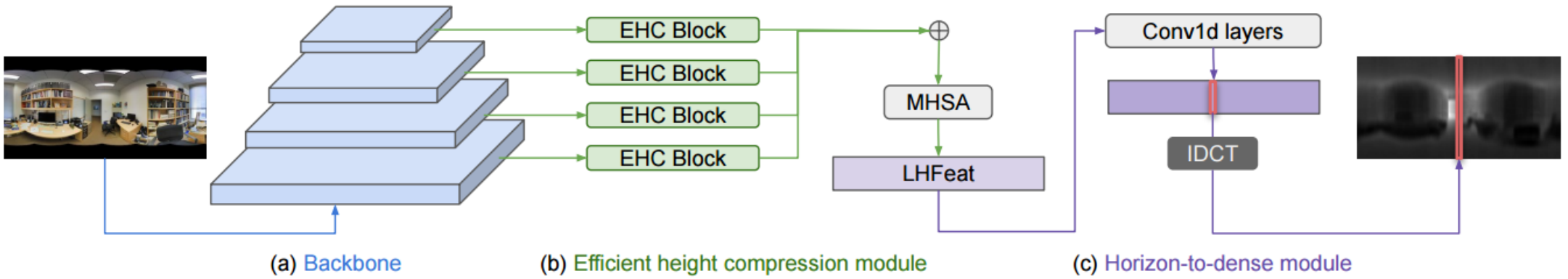

의미론적 분할은 장면 모델링의 기본 작업이다.

DistConv는 ERP 이미지의 조밀한 깊이 및 의미 예측을 위한 왜곡 인식 변형 가능한 컨볼루션 레이어를 제안한다.
360 의미 있는 분할을 위한 최근의 대부분의 방법은 정이십면체 mesh와 관련된 표현으로 작동하는 훈련 가능한 층을 설계한다.
그러나 위의 모든 방법은 파노라마 신호에 대해 비교적 낮은 해상도로 실행된다.

탄젠트 이미지는 고해상도 파노라마를 처리하고 미리 훈련된 가중치를 투시 이미지에 배치할 수 있는 세분화된 20면체에 접하는 다중 평면 이미지에 전방위 신호를 투사한다.
탄젠트 이미지와 마찬가지로 HoHoNet도 고해상도 이미지에서 작동할 수 있으며, 이는 더 나은 의미 있는 분할 정확도를 달성하는 데 필수적인 요소로 나타났다.

최근의 방법과 대조적으로 HoHoNet은 ERP 이미지에서 직접 실행되며 고도로 최적화된 딥 러닝 라이브러리는 모든 작업을 쉽게 구현할 수 있다.

# 3 HoHoNet - 조밀한 깊이 추정을 위한 HoHoNet 프레임워크 개요 ▶ ❶ 전체 구성




① 고해상도 파노라마는 먼저 backbone(예: ResNet)에 의해 처리된다.

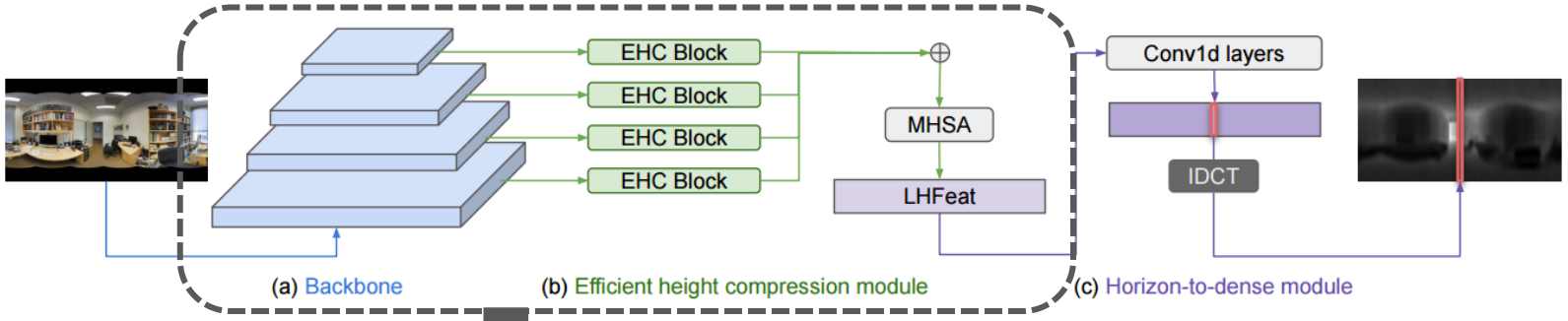
② 형상 피라미드는 제안된 EHC(Efficient Height Compression) 모듈과 정교화를 위한 다중 헤드 자기 주의(MHSA) 모듈에 의해 압착 및 융합된다.

그 결과 LHFeat은 compact며(예: 입력 이미지가 R 3×512×1024인 경우 R 256×1024), 전체 네트워크가 기존의 인코더-디코더 네트워크보다 훨씬 빠르게 조밀한 기능을 실행할 수 있다는 점에 유의한다.

③ 마지막으로, 최종 예측을 산출하기 위해 1D 컨볼루션 레이어를 사용한다.

우리는 DCT 주파수 영역에서 예측이 우수한 결과를 가져온다는 것을 발견하여 각 열의 예측에 IDCT를 적용한다.

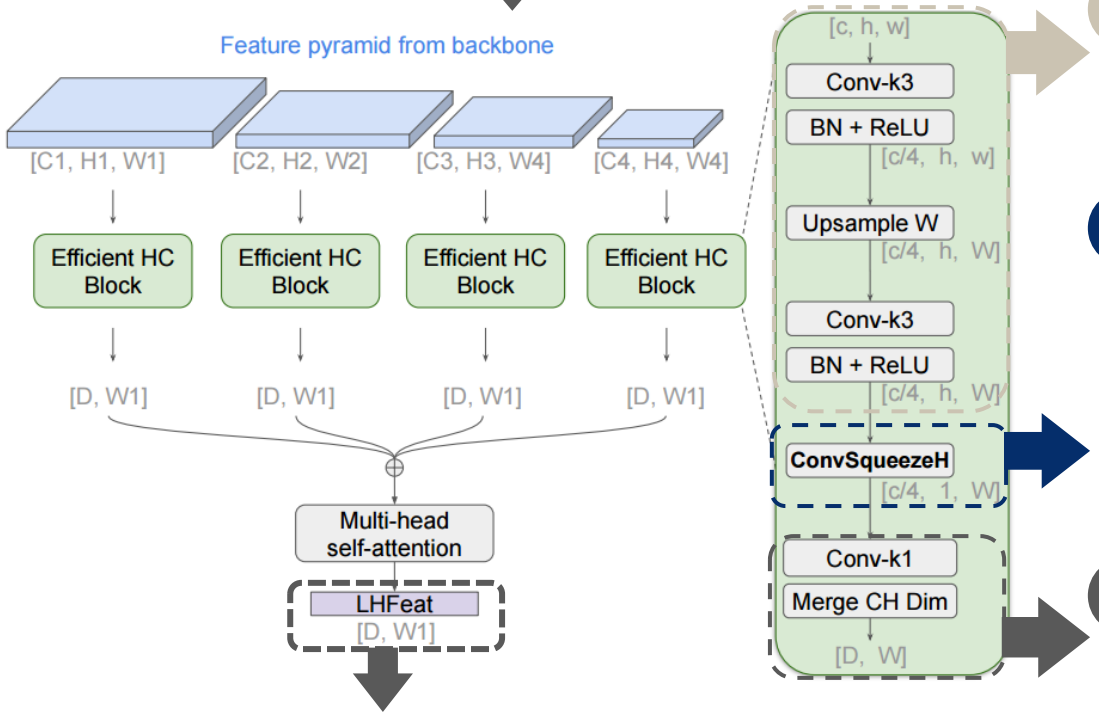
# 3 HoHoNet - 조밀한 깊이 추정을 위한 HoHoNet 프레임워크 개요 ▶ ❷ LHFeat에 대한 EHC(효율 높이 압축) 모듈




먼저 backbone의 피라미드에서 각 2D feature의 높이를 짜내기 위해 EHC 블록을 사용한다.

그 결과 1D 형상이 합산으로 간단히 융합된다.

2D 및 1D 형상의 크기는 각각 [C, H, W] 및 [C, W]로 표기한다.







HorizonNet의 LSTM = HoHoNet의 LHFeat에 대한 EHC(효율 높이 압축) 모듈

1

EHC 블록 내에서 입력 2D feature은 채널 감소를 위해 먼저 Conv2D 블록에 의해 처리된 다음,

필요한 경우 공간 너비가 $W_1$로 upsampling되고 마지막으로 다른 Conv2D 블록이 upsampling된 feature을 개선한다.

2

feature 높이를 1로 효율적으로 줄이기 위해 커널 크기를 (h, 1)로 설정하여 패딩 없이 전체 feature 높이를 커버하는 깊이 있는 컨볼루션 레이어인 ConvSqueeZH 레이어를 설계한다.

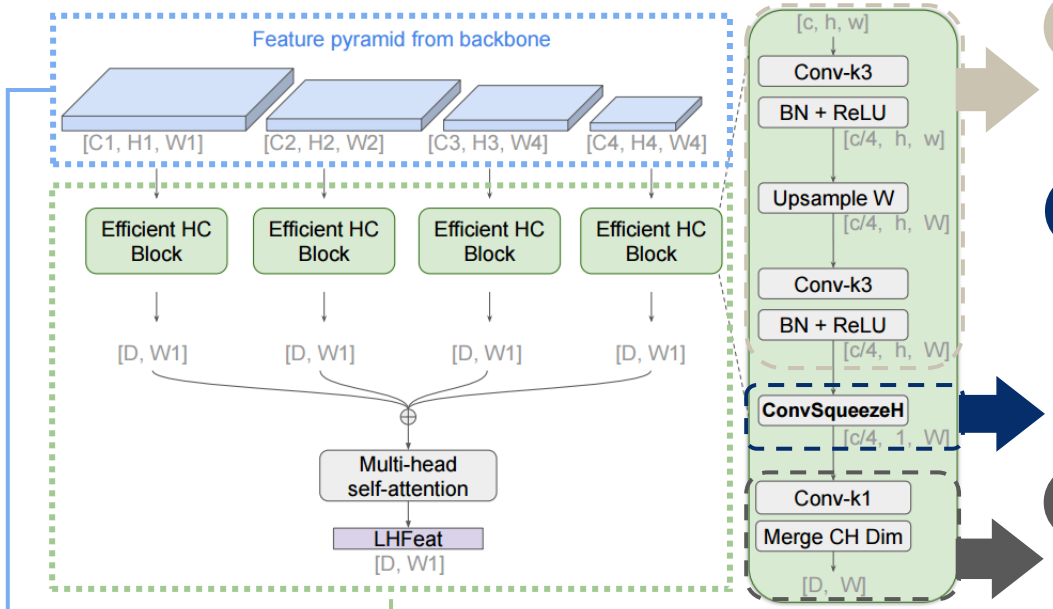
ConvSqueezeH 레이어는 커널 크기가 패딩 없이 이전에 알려진 입력 형상 높이로 설정된 깊이 별 컨볼루션 레이어로 출력 형상 높이 1을 생성한다.

각 EHC 블록의 파라미터 h는 $H_{inp}$가 주어지면 자동으로 사전 계산된다.

3

마지막으로, Conv2D 레이어는 채널 수를 LHFeat의 잠재 크기 D로 변환하고, 높이 치수는 ConvSqueezH 레이어에 의해 이미 1로 감소되었기 때문에 간단히 폐기된다.

# 3 HoHoNet - 조밀한 깊이 추정을 위한 HoHoNet 프레임워크 개요 ▶ ❷ LHFeat에 대한 EHC(효율 높이 압축) 모듈







1

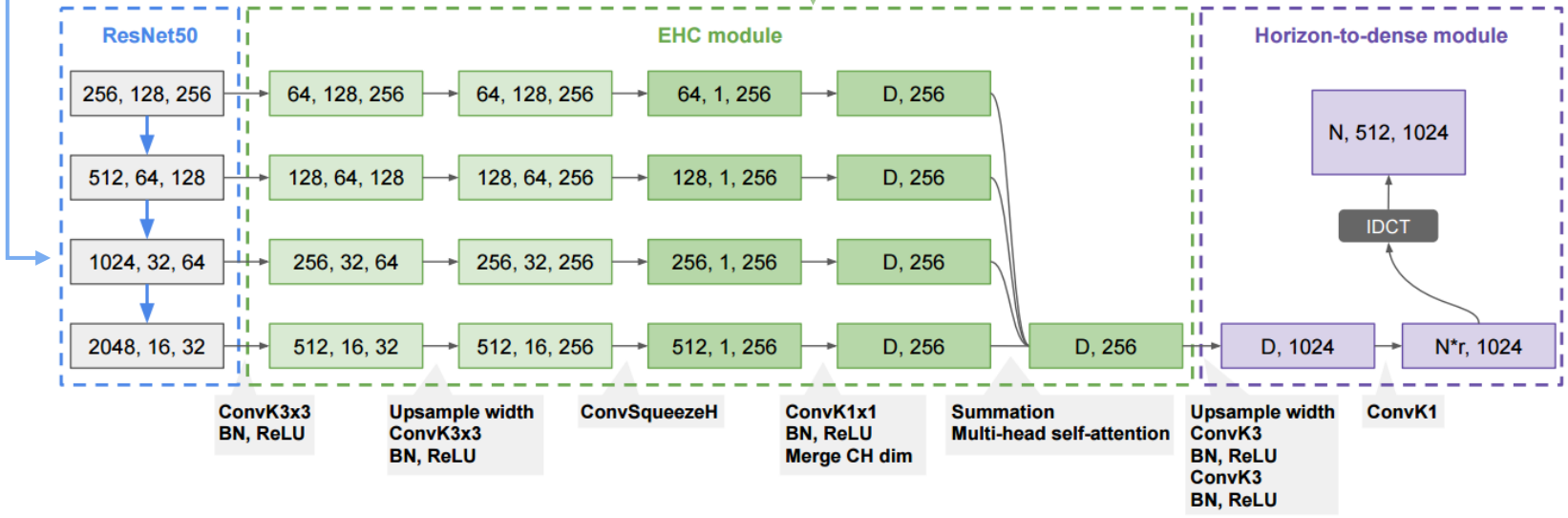
EHC 블록 내에서 입력 2D feature은 채널 감소를 위해 먼저 Conv2D 블록에 의해 처리된 다음,
필요한 경우 공간 너비가 $W_1$로 upsampling되고 마지막으로 다른 Conv2D 블록이 upsampling된 feature을 개선한다.

2

feature 높이를 1로 효율적으로 줄이기 위해 커널 크기를 (h, 1)로 설정하여 패딩 없이 전체 feature 높이를 커버하는 깊이 있는 컨볼루션 레이어인 ConvSqueeZH 레이어를 설계한다.
ConvSqueezeH 레이어는 커널 크기가 패딩 없이 이전에 알려진 입력 형상 높이로 설정된 깊이 별 컨볼루션 레이어로 출력 형상 높이 1을 생성한다.
각 EHC 블록의 파라미터 h는 $H_{inp}$가 주어지면 자동으로 사전 계산된다.

3

마지막으로, Conv2D 레이어는 채널 수를 LHFeat의 잠재 크기 D로 변환하고, 높이 치수는 ConvSqueezH 레이어에 의해 이미 1로 감소되었기 때문에 간단히 폐기된다.




입력 파노라마의 높이와 너비는 각각 512와 1024로 가정한다.

D와 E는 하이퍼 파라미터다.

# 3 HoHoNet - 조밀한 깊이 추정을 위한 HoHoNet 프레임워크 개요 ▶ ❷ LHFeat에 대한 EHC(효율 높이 압축) 모듈

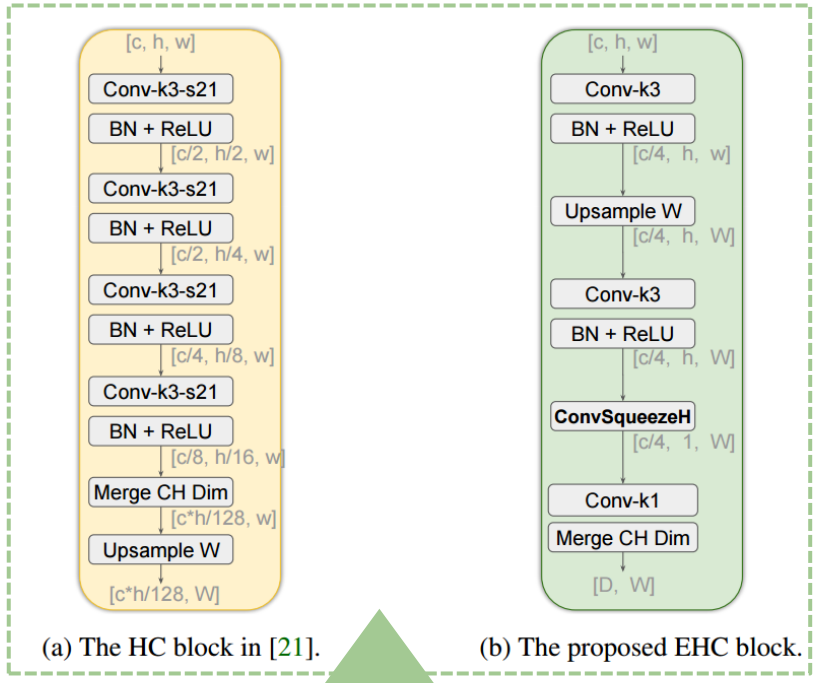
**EHC 블록과 HC 블록 비교**

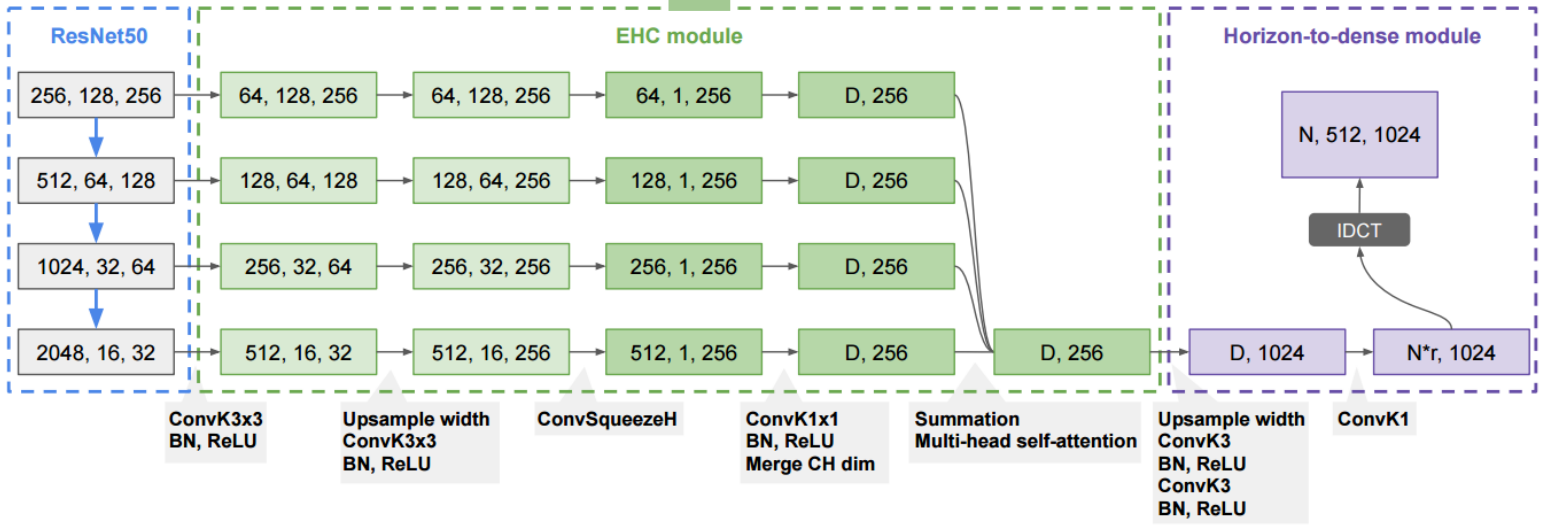
높이 압축 블록은 백본에서 2D 형상을 압착하여 1D 수평 형상을 생성하는 것을 목표로 한다.

HC 블록은 일련의 컨볼루션 레이어를 사용하여 채널 수와 높이를 점차적으로 감소시킨다.

반면, EHC 블록은 먼저 채널 감소를 위한 컨볼루션 레이어를 사용한 다음 이중선형 업샘플링 및 ConvSqueezeH 레이어를 사용하여 수평 형상의 형상을 생성한다.

절제 실험에서 HC 블록을 제안된 ECH 블록으로 대체하면 속도와 정확도가 향상된다.

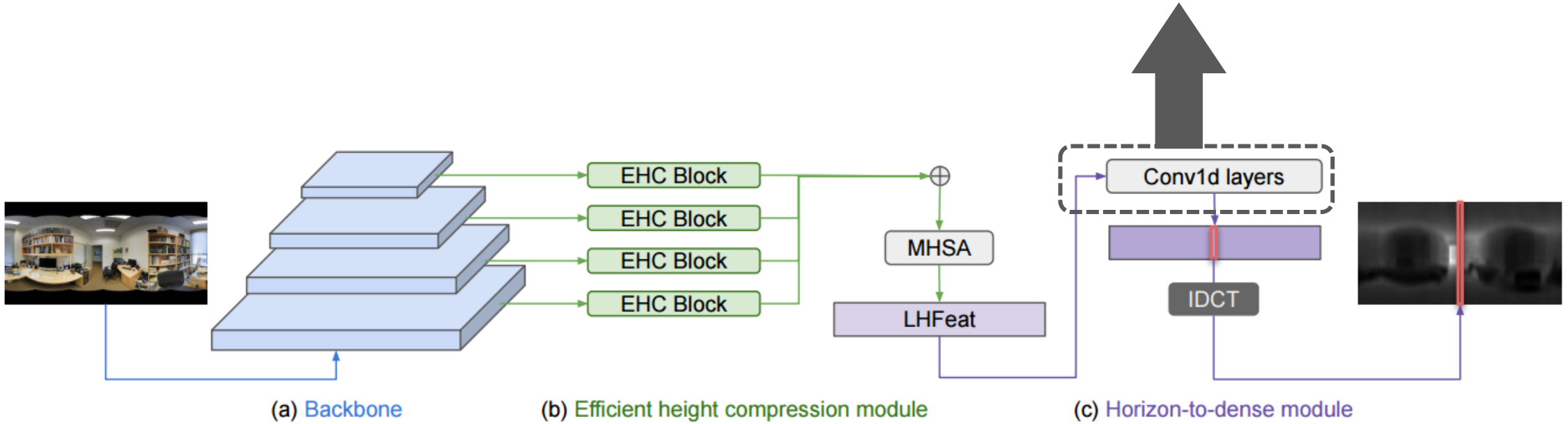


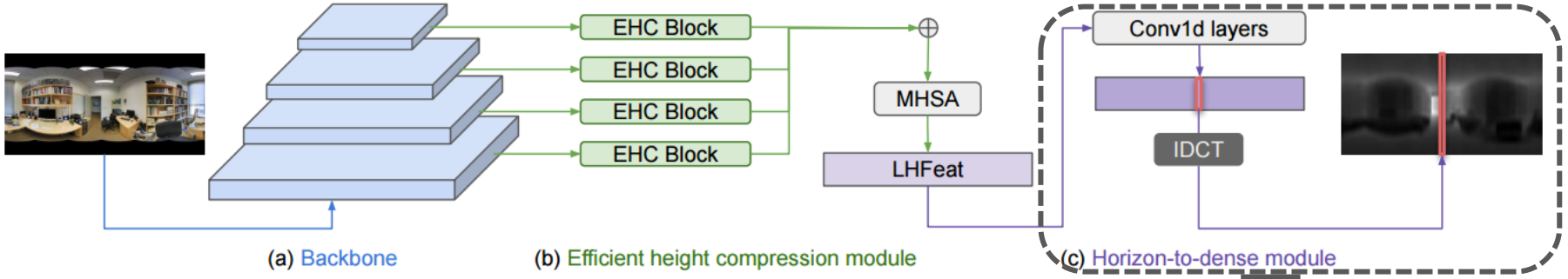
# 3 HoHoNet - 조밀한 깊이 추정을 위한 HoHoNet 프레임워크 개요 ▶ ❸ horizon-to-dense 모듈, 열 당 1D 양식 예측

1D 양식을 예측하기 위해 먼저 $R^{D \times W_1}$에서 $R^{D \times W_{inp}}$ 로 수평 형상을 upsampling하고 BN, ReLU 사이에 커널 크기 3, 3, 1의 Conv1D 레이어를 각각 적용한다.

마지막 레이어는 $R^{D \times W_{inp}}$ 에서 최종 예측을 산출한다.



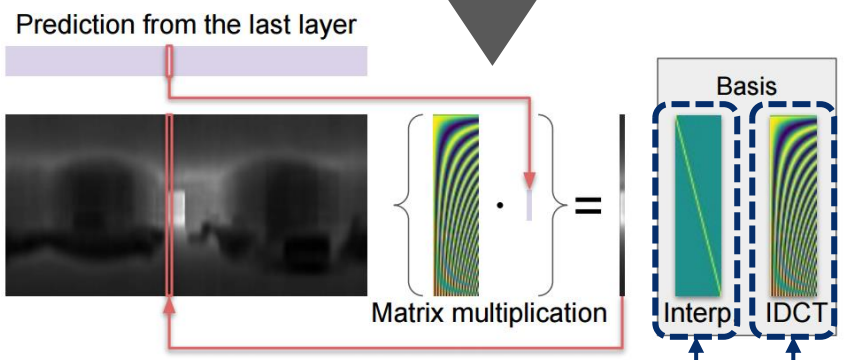
# 3 HoHoNet - 조밀한 깊이 추정을 위한 HoHoNet 프레임워크 개요 ▶ ❸ horizon-to-dense 모듈, 픽셀당 2D 양식 예측




출력 공간을 열당 형식으로 shaping하는 전략은 픽셀당 양식이 포함된 작업에는 적용되지 않는다.

여기서는 compact LHFeat $R^{D\times W_1}$에서 조밀한 예측 $R^{N\times H_{inp}\times W_{inp}}$를 도출하기 위한 HoHoNet의 수평 대 밀도 모듈을 제시한다.

이 기능은 다양한 애플리케이션에 보다 일반적인 시나리오의 문을 열어준다.




2D 양식 예측을 위한 훈련 가능한 계층은 출력 계층의 채널 수가 $E = N \cdot r$로 증강되고 여기서 $N$은 작업에 대한 대상 채널의 수이고 $r$은 이미지 열에 의해 공유되는 구성 요소의 수라는 점을 제외하면 3.3항에서 소개한 1D 예측을 위한 계층과 거의 동일하다.

생성된 예측은 $R^{E\times W_{inp}}$ 에서 $R^{N\times r\times W_{inp}}$ 로 재구성된다. 예측된 r 값에 할당한 물리적 의미에 따라 각 열에 대해 $R^r$를 $R^{H_{inp}}$로 복구하기 위한 두 가지 다른 연산을 제시한다.

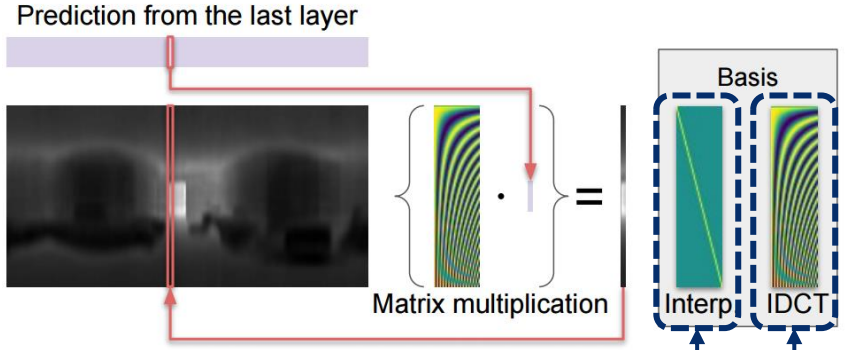
① 보간법 (interpolation)

② 역 이산 코사인 변환 (IDCT, Inverse Discrete Cosine Transform)

# 3 HoHoNet - 조밀한 깊이 추정을 위한 HoHoNet 프레임워크 개요 ▶ ❸ horizon-to-dense 모듈, 픽셀당 2D 양식 예측

제안된 수평 대 밀도(h2d) 모듈은 compact LHFeat에서 밀도 예측을 생성할 수 있다.

Linear interpolation을 IDCT로 대체함으로써 조밀한 예측 결과를 개선할 수 있다.
수평 대 밀도 모듈(h2d)을 통해 효율적으로 인코딩된 LHFeat은 이제 조밀한 양식을 모델링할 수 있다.




각 열의 예측은 기초 M의 성분들의 선형 조합에 대한 가중치로 작용한다.

① HoHoNet은 M이 선형 보간을 구현하는 경우 공간 영역에서 예측하고

② M이 IDCT를 구현하는 경우 주파수 영역에서 학습한다.

**① 보간법 (interpolation)**

가장 간단한 방법은 잠재 치수 $r$을 출력 높이로 보고 선형 보간법을 적용하여 $r < H_{inp}$일 경우, $H_{inp}$의 $r$ 크기를 조정하는 것이다.

**② 역 이산 코사인 변환 (IDCT, Inverse Discrete Cosine Transform)**

에너지 압축 특성에 대한 이미지 압축에서 DCT의 적용에 영감을 받아, r 예측 값을 높은 주파수가 잘리는 DCT 주파수 영역에 있는 것처럼 본다.
이 경우 IDCT를 적용하여 low-pass 신호를 원래 신호로 복구할 수 있다.

IDCT가 선형 보간법을 지속적으로 능가한다.

LHFeat은 공간-행 정보를 혼합하므로, 평평한 행이 없는 LHFeat에서 행에 의존하는 밀도 양식을 분리하기 위해 마지막 층을 훈련시키는 것은 문제가 될 것이다.

반대로, 주파수 영역에서 예측하는 법을 배우는 것은 각 열의 원래 행 정보를 전체적으로 특징짓는 의미 있는 공간 주파수를 가진 잘 정의된 기본 함수로부터 이익을 얻을 수 있으므로 행 의존성 문제를 완화시킬 수 있다.

# 4 비교 정리

# ENTER THE CONTENTS

I believe that someone like you. I was broken my heart. But now, I am standing again.

### Someone

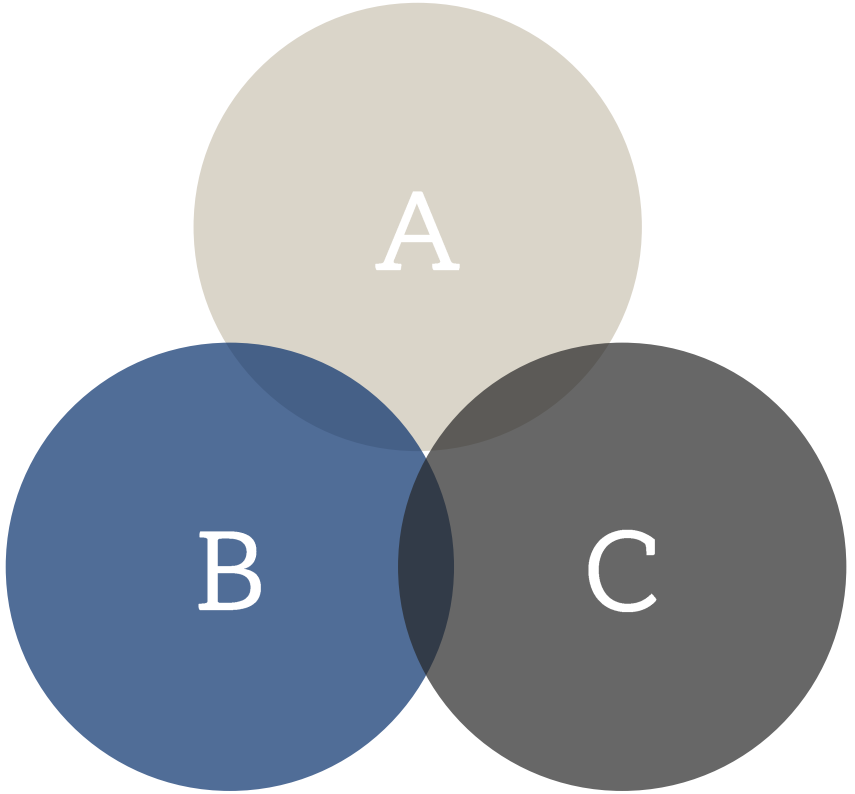
I believe that someone like you. Enter something here.

### Someone

I believe that someone like you. Enter something here.

### Someone

I believe that someone like you. Enter something here.



# 1 CONTEN TS

## ENTER THE CONTENTS

I believe that someone like you. I was broken my heart. But now, I am standing again.


### Something

I believe that someone like you. Enter something here.

### Something

I believe that someone like you. Enter something here.

### Something

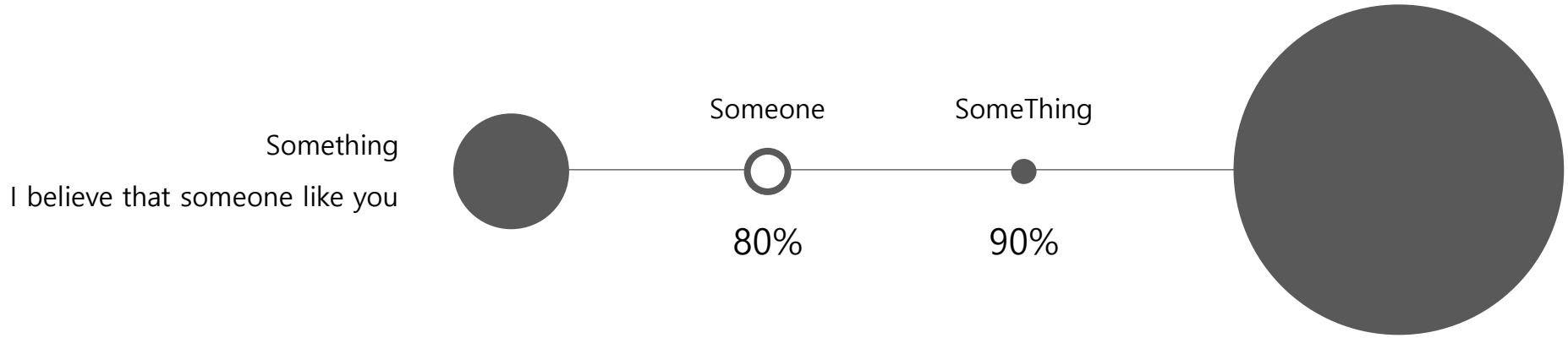
I believe that someone like you. Enter something here.

# 2 CONTENTS

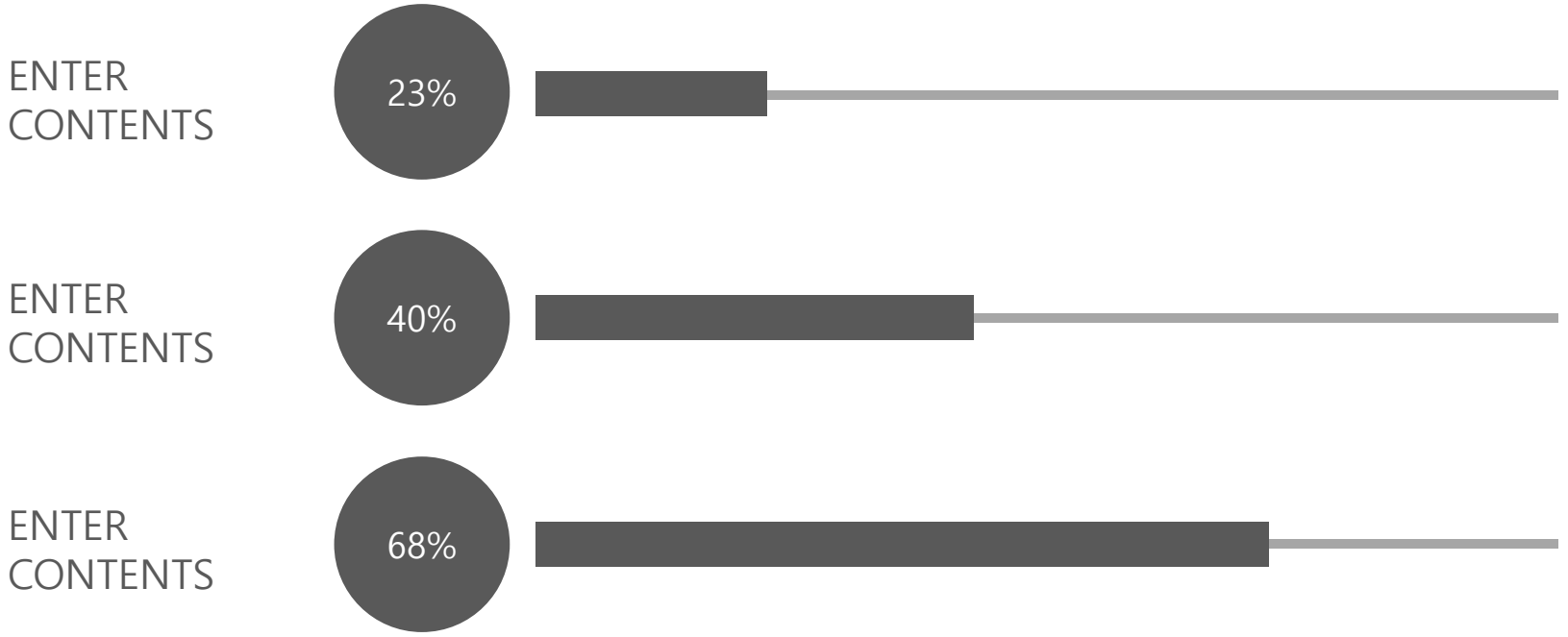
ENTER THE CONTENTS

I believe that someone like you

ENTER



# 3 CONTENTS




I want you to use this template for free and to remember slug and CREBUGS for me.

# 4 CONTENTS

## ENTER THE CONTENTS

I believe that someone like you. I was broken my heart. But now, I am standing again.

### Someone

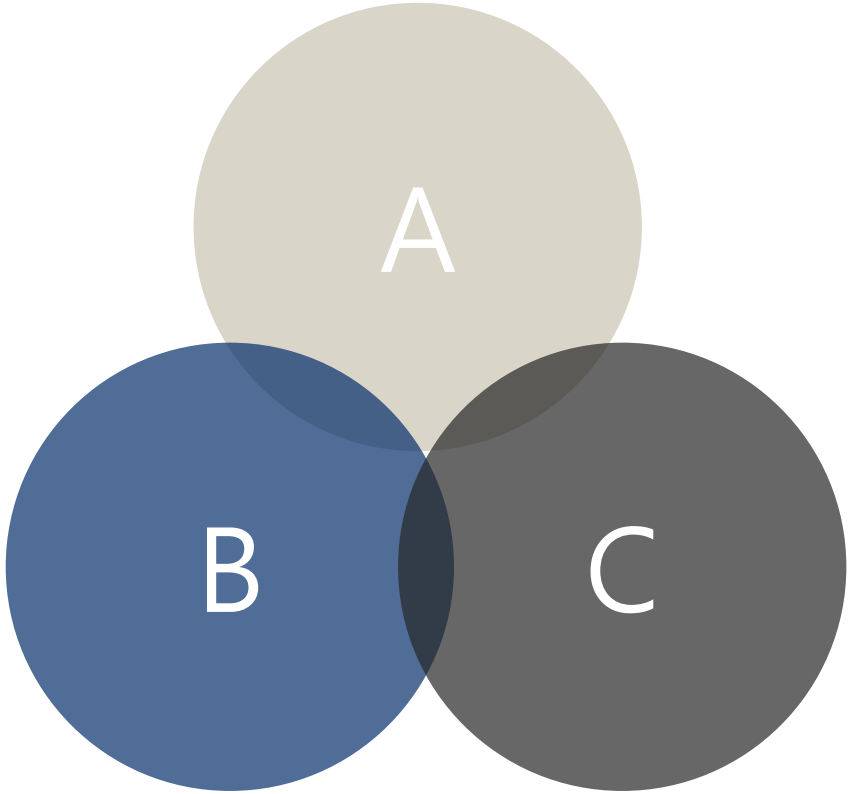
I believe that someone like you.
Enter something here.

### Someone

I believe that someone like you.
Enter something here.

### Someone

I believe that someone like you.
Enter something here.



# 5 CONTENTS

ENTER THE CONTENTS

I believe that someone like you. I like pizza.
I already know that you want to meet me.


ENTER THE CONTENTS

I believe that someone like you. I like pizza.
I already know that you want to meet me.


ENTER THE CONTENTS

I believe that someone like you. I like pizza.
I already know that you want to meet me.


ENTER THE CONTENTS

I believe that someone like you. I like pizza.
I already know that you want to meet me.

# THANK YOU SO MUCH!