IEEE International Conference on Image Processing
25-28 October 2020, United Arab Emirates

Does Super-Resolution improve OCR performance in the real world? A case study on images of receipts

TEC-01 -- Interpolation, Super-Resolution and Mosaicing

Vivien Robert
Spendesk, former Artefact
Hugues Talbot
Universite Paris-Saclay, CentraleSupelec, Inria

25th October 2020

Outline

1. Outline
2. Introduction
3. Single Image Super-Resolution business use case
4. Results
5. Conclusion

Introduction

There are some issues with the current single image super-resolution (SISR) approaches:

- Mainly focus on improving the perceptual image quality from a **human** point of view, it does not translate well when used to enhance another computer vision task;
- SISR is usually trained and evaluated on synthetic datasets. This may be not realistic in **real-world applications;**
- The generated images can suffer from the creation of unwanted patterns and hallucinating artifacts which can lead to **misinterpretation and errors**.

No real evidence that SISR can improve another computer vision task in **real business cases of computer vision application**.

We experiment using SISR into a real business use case and provide a practical point of view on it

- **Loyalty program** developed by a beverage company
- Consumers are offered rewards when they **send receipts of company's product purchases**, either in supermarkets or bars

OR

1 The customer buys company's product and get a receipt

2 The customer uploads it on Messenger chatbot

GOOGLE OCR

3 OCR detects the text and useful information is extracted (date, products…)

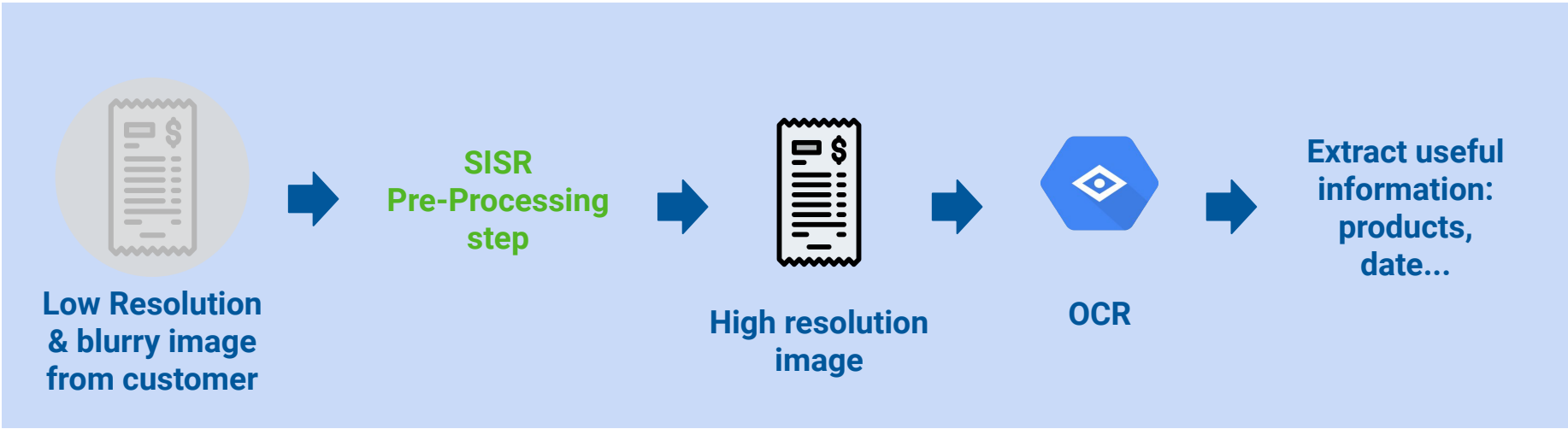
4 The customer receives credits proportionally to the amount paid

OCR alone fails to extract the information on challenging images of receipt

*Among the readable images, OCR fails to extract the relevant information in **30%** of the cases*

-  Handheld and telephone quality photographs with **poorly printed text**
- **Numerous artifacts** (blur, flash reflections, shadows, malfunctioning autofocus) as the photographs are not professionally taken in bars
-  Photographs taken **at a large distance** from the receipts to take the whole receipt in the same picture
-  **Automatic compression** during peak hours can lower picture resolutions

SISR is used as pre-processing step to help OCR better extract information



The quality of the training set is key to build a successful SISR model

The data generation process should be as close as possible to the reality

1

Images as close as possible to the business case

- ➔ Images from real customers already using the loyalty program
- ➔ Select those already well detected by the OCR
- ➔ Increase data size with images from other sources (ICDAR 2019)

2

Custom degradation function which mimics the use case reality

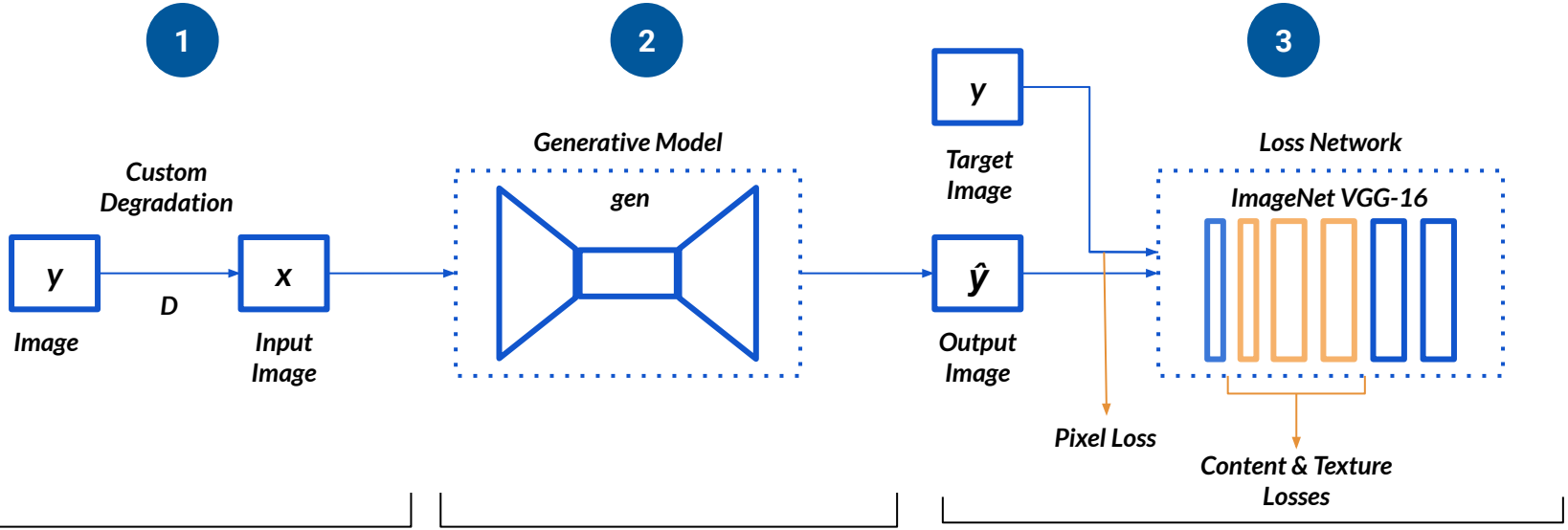
- ➔ Bilinear interpolation
- ➔ Compress the image using the JPEG lossy standard coding
- ➔ A Gaussian Blur

3

Generate pairs of (LR, HR) images for the supervised learning

- ➔ Apply the custom degradation function to the HR images

We train a Deep Learning model to enhance the resolution of input images



Custom Degradation

- Input: HR target images
- Output: LR degraded images

Generative Model

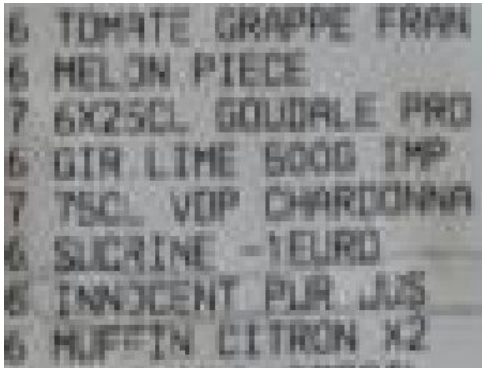
- Input: LR images
- Output: HR cleaned images
- Model: Resnet 34 UNET

Perceptual Loss

- Inputs: HR cleaned images & HR target images
- Model: ImageNet Pre-trained VGG-16

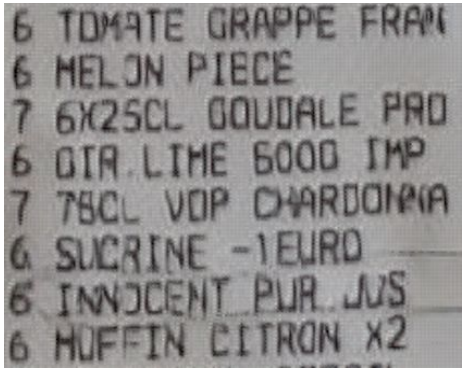
Visually, the quality of images is improved on challenging images

Before SISR



No line perfectly recognized by the OCR

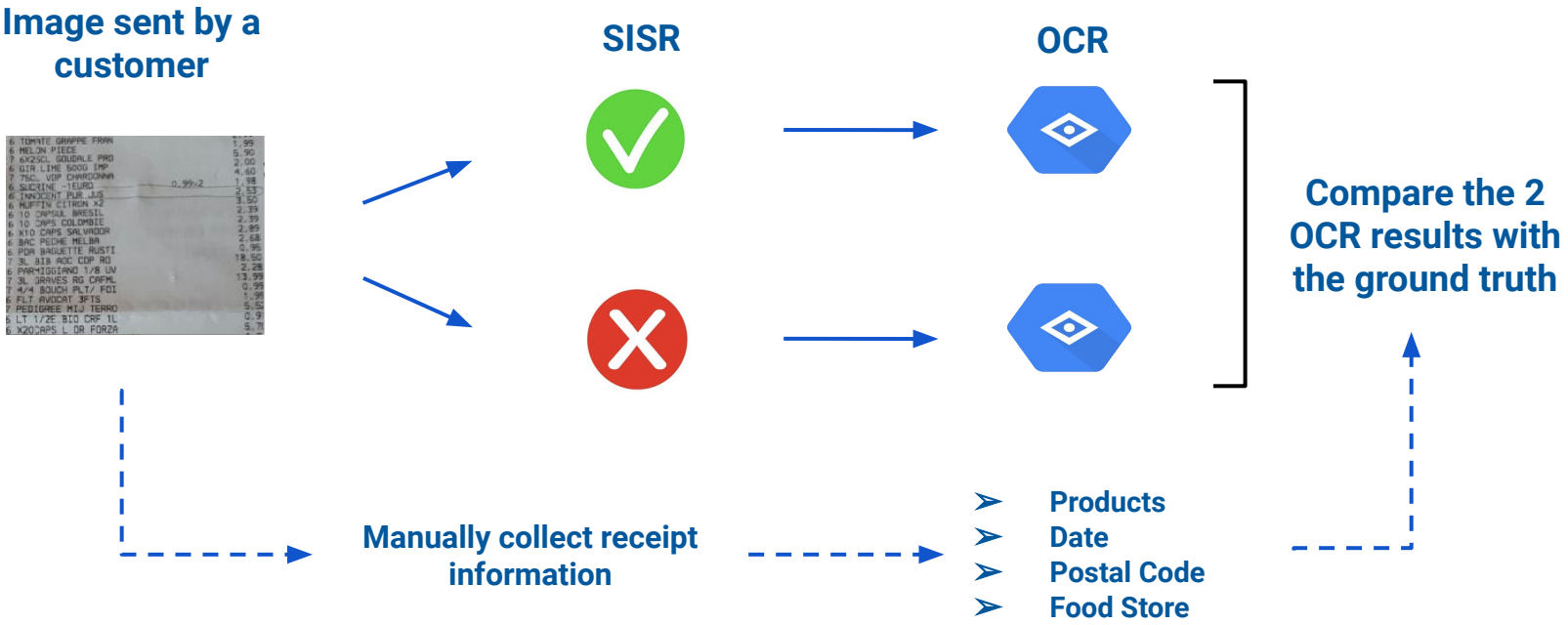
After SISR



5 lines perfectly recognized by the OCR

We use a task-based evaluation method to objectively evaluate the model performance

Compare the OCR performance with and without the SISR preprocessing step



To measure the benefits, we compute the increase in detection of the relevant items on challenging images

Scope

*Challenging receipt images, i.e images where OCR **alone performed poorly,** failed to detect the relevant information*

With SISR

+15% postal code

+10% detected products

+7% food store name

+5% receipt date

The SISR model **improves by up to 15%** the extraction of relevant information

What about non-challenging images already well detected by the OCR alone?

Scope

Non-challenging receipt images, i.e images where OCR alone managed to detect the relevant information

With SISR

-9% postal code

-6% food store name

-5% detected products

-1% receipt date

The detection is **damaged by up to -9%**

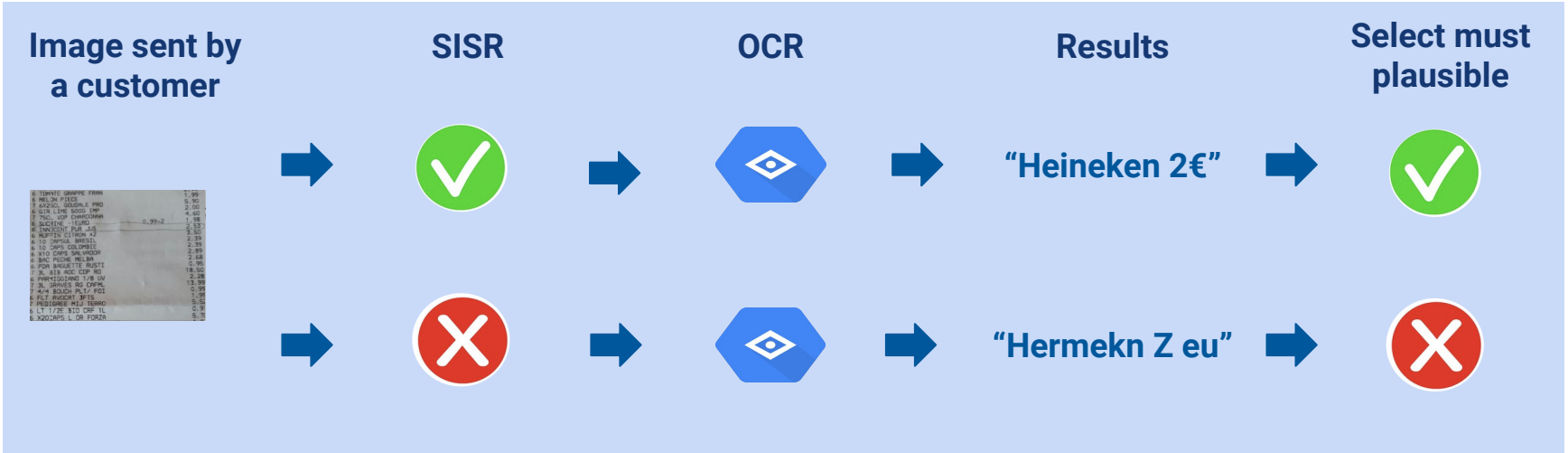
Why?

- Creation of pattern distortions
- Information loss due to inevitable reshape

How can we leverage the SIRS ability to enhance OCR performance on challenging images while not damaging it on HR ones?

For real-time production, we advise to build 2 tracks, one with and one without SISR

- **run in parallel OCR both with and without the SISR model**
- **keep only the most plausible result from the two runs** (i.e. with correct format of date, real product names, etc)



Conclusion

We evaluate the influence of Single Image Super-Resolution on OCR performance in challenging circumstances

➔ SISR can improve more generalist models such as Google OCR in **specific business cases** if you have:

- A sufficient amount of data extracted directly from the use case;
- A custom degradation function close enough to the reality;

➔ SISR can be counter-productive in certain cases. For real-time production purpose, we recommend to **build 2 tracks** (with and without SISR) :

- Assess the image quality first and then put it in the right track
- Run both with and without SISR and keep the most relevant result

Thank you!