תאריך: ‏21/11/2020

מסמך אפיון פרויקט

|  |  |  |
| --- | --- | --- |
| **הפרויקט** | | |
| שם הפרויקט: | Seeing Sound: Estimating Image from Audio | |
| מס' ב-LabAdmin: | 5952 | |
| סמסטר: | חורף 20/21 | |
| חד/דו סמסטריאלי: | חד סמסטריאלי | |
| **הצוות** | | |
| שם המנחה: | תמר רוט – שחם, עידן קליגוסר | |
| שם סטודנט 1: | יהב וינוקור | מקצוע רישום: פרויקט א' |
| שם סטודנט 2: | שגיא גרש | מקצוע רישום: פרויקט א' |
| **☐ חברה מלווה - אין** | | |
| שם החברה: | - | |
| שם איש קשר: | - | |

1. **מטרת הפרויקט**

מטרת הפרויקט היא לאמן רשת לומדת כך שתדע לקחת אות שמע כקלט, ולהוציא כפלט תמונה משוערכת של המקור ממנו נוצר אות השמע.

1. **פירוט הנחות ודרישות**

הנחות בפרויקט:

* קיימות רשתות שיודעות לייצר תמונות "מומצאות" מקוד/רעש. (auto encoders/GANs)
* קיימות רשתות אשר יודעות לקבל כקלט אות שמע של אדם מדבר ולשערך תמונה של האדם המדבר.
* קיימות רשתות לסיווג אותות קוליים.

דרישות על התוצאות:

* בפרויקט נרצה להמשיך ולשלב עיבוד אותות, קול ותמונות ולהרחיב את המימוש לתמונות כלשהן (לא בהכרח פנים).

1. **פתרונות אפשריים וסיכום קצר של סקר הספרות**

סקר הספרות שלנו התמקד במאמר Speech2Face: Learning the Face Behind a Voice.

במאמר מוצגת מערכת רשתות אשר מקבלת אות דיבור ומשערכת תמונה של האדם המדבר.

מערכת הרשתות אשר מוצגת במאמר משתמשת בשלוש רשתות:

1. רשת לזיהוי פנים – מערכת קיימת ומאומנת מראש.

2. רשת S2F – מקבלת אות דיבור ומוציאה וקטור 4096D בדומה לרשת זיהוי הפנים.

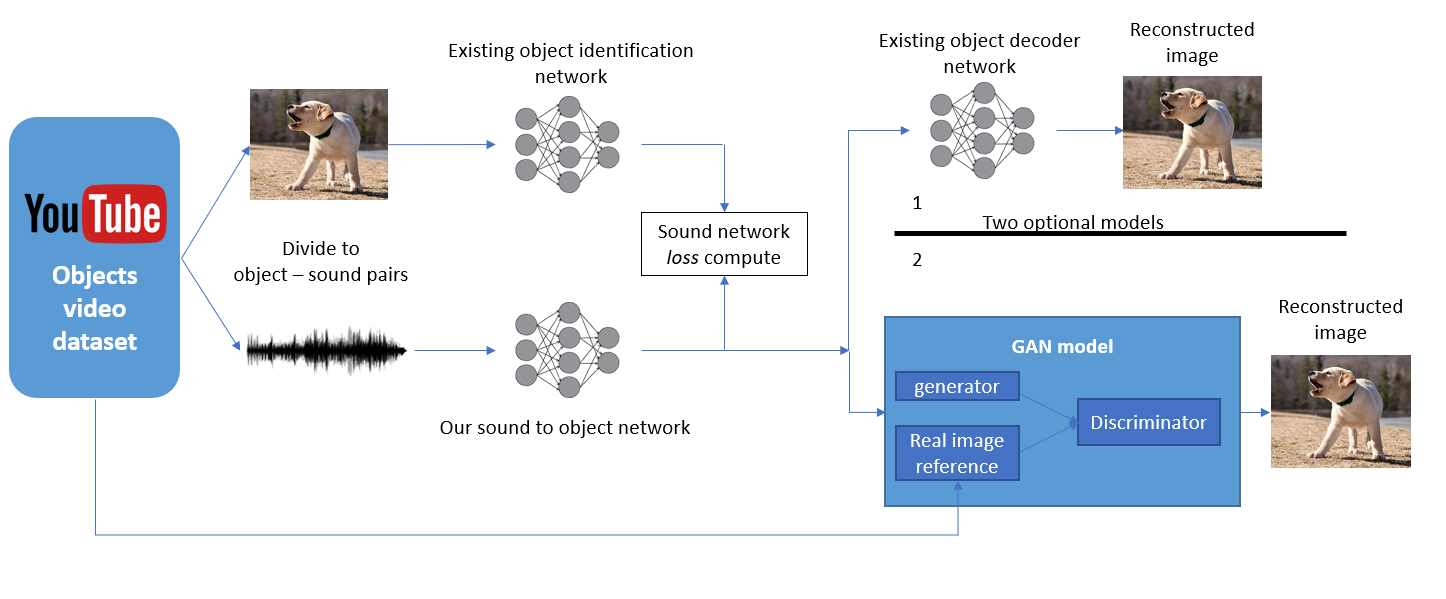
3. רשת פענוח – גם כן מערכת קיימת ומאומנת מראש, אשר יודעת לקחת וקטור פיצ'רים של 4096D ולייצר ממנו תמונה של פנים.

רשת הS2F לומדת על ידי השוואה, של הפלט שלה לתוצאות הרשת לזיהוי פנים, להוציא כפלט וקטור פיצ'רים אשר מתאים ליצירת תמונה על ידי רשת הפענוח.

באופן דומה נרצה אנחנו להשתמש במערכת קיימת לזיהוי אובייקטים בתמונה (כגון VGG16/19) ולאמן רשת לניתוח אותות קוליים אשר תלמד להוציא וקטור פיצ'רים מתאים לתמונה שממנה הגיעה אות הקול (אותו וקטור שהוציאה הרשת לזיהוי אובייקטים) ולשחזר תמונה.

שיחזור התמונה יכול להתבצע באמצעות 2 אפשרויות:

1. רשת מפענח, בדומה למוצג במאמר, אשר מקבלת וקטור פיצ'רים ויודעת להחזיר תמונה.
2. שימוש בGAN במקום מפענח. אשר תקבל תמונות מה-dataset ותייצר תמונה מתאימה – לפי אות הקול. המערכת תלמד, באמצעות ה descriminator, לייצר תמונות שמתאימות ל-dataset.
3. **תרשים מלבנים (block diagram) של הפתרון הנבחר או הנבדק**



בתרשים ניתן לראות את 2 האופציות למימוש שהצגנו בסעיף פתרונות אפשריים:

ראשית, בשני הפתרונות קיימות 2 רשתות – האחת לזיהוי אובייקטים בתמונה למשל VGG19). השנייה לניתוח שמע, אותה נאמן להוציא כפלט, וקטור פיצ'רים אשר מתאים לוקטור הפיצ'רים שהוציאה הרשת לזיהוי האובייקטים.

לאחר מכן קיימות 2 אופציות לשלב יצירת התמונה -

הראשונה, שימוש ברשת אשר יודעת לקבל וקטור פיצ'רים ולהוציא כפלט תמונה. (למשל רשת מסוג auto encoder)

השנייה, שימוש ב- GAN, אשר תקבל בתור "רפרנס" תמונות מה Dataset, ובעזרת וקטור הפיצ'רים שיצרה הרשת לניתוח קול, תדע לייצר תמונה שנראית אמיתית, ומתאימה לסיטואציה אשר נשמעה באות הקולי.

1. **מודולים שנידרש לפתח**

נדרש לפתח ולאמן את הרשת לזיהוי אובייקטים על ידי אות שמע – אשר נרצה שתוציא וקטור פיצ'רים מתאים לוקטור הפיצ'רים שהוציאה הרשת לזיהוי אובייקטים.

בנוסף נרצה לפתח GAN מתאים אשר בהינתן הוקטור הנ"ל ידע לייצר תמונה מתאימה.

1. **מודולים מוכנים שניתן להיעזר בהם**

נוכל להשתמש ברשת לזיהוי אובייקטים קיימת כגון VGG16/19.

קיימות רשתות GAN אשר נוכל להשתמש בהן/ לשנות אותן לצרכינו, כגון BigGAN.

1. **סביבת עבודה וכלי פיתוח שיהיו בשימוש**

PyTorch

1. **שיטת הבדיקה שתידרש בסיום הפרויקט**

בשלב הראשון נוכל להשתמש ברשת לזיהוי אובייקטים, על מנת לקבל הערכה של טיב התוצאות.

1. **רשימת משימות:**

|  |  |  |
| --- | --- | --- |
| מס' | שם המשימה | תיאור המשימה |
| 1 | רכישת ידע וקריאת מאמרים | קורס Stanford CS231n, מאמר Speech2Face. |
| 2 | מימוש מסווג אודיו | מימוש רשת מסווגת אודיו. |
| 3 | בחינת פתרונות אפשריים לרשת המאומנת לקידוד ופיענוח קוד לתמונות (auto-encoder/GANS) |  |
| 4 | בחירת הרשת האחראית למסלול קידוד ויצירת התמונה |  |
| 5 | הרכבת dataset מתאים לאימון | למשל - מתוך סרטוני youtube. |
| 6 | אימון encoder קולי אל מול המסלול המאומן של התמונות |  |
| 7 | בחינת התוצאות ואבלואציה |  |
| 8 |  |  |
| 9 |  |  |
| 10 |  |  |

1. **תרשים גאנט (התקדמות הפרויקט):**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| מספר חודשים מתחילת הסמסטר | | | | | | | | | משימה |
| 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 |  |
|  |  |  |  |  |  |  |  |  | 1 |
|  |  |  |  |  |  |  |  |  | 2 |
|  |  |  |  |  |  |  |  |  | 3 |
|  |  |  |  |  |  |  |  |  | 4 |
|  |  |  |  |  |  |  |  |  | 5 |
|  |  |  |  |  |  |  |  |  | 6 |
|  |  |  |  |  |  |  |  |  | 7 |
|  |  |  |  |  |  |  |  |  | 8 |
|  |  |  |  |  |  |  |  |  | 9 |
|  |  |  |  |  |  |  |  |  | 10 |