לילה טוב חברים, היום אנחנו שוב בפינתנו DeepNightLearners עם סקירה של מאמר בתחום הלמידה העמוקה היום בחרתי לסקירה את המאמר שנקרא:
Contrastive Learning Of Medical Visual Representations From Paired Images And Text
שיצא לפני שבוע וחצי.

תחום מאמר: למידת ייצוג (representation learning) לצילומים רפואיים

מאמר הוצג בכנס: טרם ידוע

תמצית מאמר: המאמר מציע שיטה לבניית ייצוג מימד נמוך של צילומים רפואיים תוך כדי שימוש בגישה הנקראת הנקראת (Noise Contrastive Estimation (NCE. החידוש שהמאמר מביא הוא שימוש ב NCE לבניית ייצוגים בשני דומיינים בו זמנית: הראשון זה ייצוג של התמונה (צילום) והשני ייצוג של כותרת (תיאור) טקסטואלי של צילום.

רעיון בסיסי: להבדיל מה NCE הרגיל המאמר מנסה לבנות ייצוגים של צילום וגם של הכותרת על שהייצוג של הצילום יהיה "קרוב יותר" (אחרי טרנספורמציה מסוימת) לייצוג של תיאור הצילום מאשר לתיאור של צילום אחר. תכונה דומה מתקיימת גם לזוג של תיאור טקסטואלי של צילום וזוגות של צילומים אחרים עם אותו התיאור (בגלל זה הם קוראים לגישה שלהם דו-כיוונית). לדעתם זה מאפשר להגיע לייצוג צילום המכיל בתוכו תכונות "סמנטיות חזקות מתיאור הצילום"

תקציר מאמר: בעולם צילומים רפואיים יצירת דאטה סטים מתויגים איכותיים הינה יקרה מאוד. רוב דאטה סטים קיימים מתוייגים הם לא גדולים שמקשה מאוד על אימון מודלים גדולים(רשתות נוירונים :)) בעלי יכולת הכללה טובה. מצד שני ניסיונות להשתמש בייצוגים מאומנים על דאטה סטים מדומיינים אחרים (כמודל pretrained וכיול על דאטה סט רפואי קטן) בדרך כלל לא מובילים לייצוגים חזקים בדומיין הצילומים הרפואיים כי יש הבדלים מהותיים אינהרנטיים בין תמונות טבעיות לבין צילומים רפואיים. מצד שני שימוש בגישות self-supervised לבניית ייצוגים בדומיין הרפואי נתקלים גם כן בקשיים עקב שינוים די קטנים של צילומים רפואיים מקלאסים (קטגוריות) שונות.

אז המאמר מציע לנצל דאטה טקסטואלי המלווה צילומים רפואיים בשביל לבנות ייצוגים עשירים יותר. הם מציעים שיטה הנקראת Contrastive VIsual Representation Learning from text - ConViRT. בליבה בניית ייצוגים "קרובים" לזוגות של צילום-תיאורו ו"הרחקת" ייצוגים של זוגות אקראיים של (צילום, תיאור). בעצם זה הכללה של NCE קלאסי למקרה קרוס-דומיין, כלומר בנייה של ייצוגים בשני דומיינים שונים בו זמנית להבדל מ NCE שעושה זאת בדומיין אחד. בעצם זה ניתן מענה לשוני קטן בין צילומים רפואיים מקלאסים שונים המקשה על שימוש ב NCE סטנדרטי לדומיין זה. כמו שמקובל הדומיינים אחרים עושים pretrain לרשת שבונה את הייצוג על כמה דאטה סטים גנריים ואז מכיילים אותו למשימת downstream (פיין טיונינג)

הסבר קצר על :NCE בשביל להבין איך עובד NCE קרו-דומיין בואו ניזכר מה זה NCE קלאסי. הנחת היסוד ב NCE אומרת שייצוג חזק בהכרח מסוגל להפריד בין הדוגמאות (דוגמאות קשורות או אותה דוגמא עם אוגמנטציה) לבין זוגות דוגמאות רנדומלית. בין השימושים של טכניקה זו אפשר להזכיר negative sampling שהשתמשו בו למשל ב- word2vec. ניתן להוכיח שעבור צורה מסוימת של NCE לוס (הנקראת InfoNCE שבה משתמשים במאמר זה) כי ככל שלוס זה קטן יותר המידע הדדי בין הדוגמא במרחב המקורי לבין ייצוגה במרחב ממימד נמוך עולה. זה כמובן מצביע על אובדן פחות אינפורמציה בין הדאטה המקורי לבין ייצוגה כלומר הייצוג יהיה פחות לוסי ויותר ומייצג את הדאטה. חשוב לציין שהאימון מתבצע במרחב הייצוג לא במרחב המקורי כלומר הלוס מחושב על הייצוגים במרחב ממימד נמוך. לוס NCE זה בעצם עושה הוא לוקח זוג דוגמאות קרובות והרבה דוגמאות רנדומליות ומנסה למקסם את המנה בין דמיון של זוג הקרוב לסכום הדמיונות בינו לבין דוגמאות רנדומליות

קרוס-דומיין NCE של ConViRt: אז במקום לבנות זוגות מאותו דומיין הם מציעים לבנות זוגות מדומיינים שונים. כלומר לוקחים צילום וחלק מהתיאור שלו (אפרט על כך עוד מעט) ובונים מזה זוג חיובי. אחר כך בונים זוגות רנדומליים של צילומים והתיאורים שלהם. כמובן משתמשים באוגמניציה של צילומים בשביל לבנות זוגות חיוביים. נגיד לוקחים צילום, עושים לו crop ובונים זוג חיובי עם חלקים שונים של תיאורו (פשוט דוגמים משפטים מתיאור הצילום באופן רנדומלי).
אחרי שבונים את הזוגות מעברים כל אחד מהם דרך הרשת שלו (אחד לצילום והשנייה לטקסט). לאחר מכן בונים מיני-באטץ' המכיל דוגמא אחת חיובית והשאר רנדומליות. את הצילום מעבירים דרך המקודד שלו (הם השתמשו ב- 50ResNet) ואת הטקסט מעבירים דרך המקודד שלו (כמו שאתם יכולים לנחש זה לא אחר אלא BERT). אחר-כך לוקחים את הפלטים של שני המקודדים האלו ומטילים אותם למרחב מאותו מימד שנוכל להשוותם (מעבירים את שניהם דרך רשת בעלת שתי שכבות - כמובן כל אחד מועבר דרך הרשת שלו). לאחר מכן מחשבים את מרחק הקוסיין בין הפלטים לכל זוג. בשלב האחרון מציבים את המרחקים האלו לשתי פונקציות לוס InfoNCE: בראשון יש המכנה מכיל את סכום (אקספוננטים) של כל המרחקים בין כל הזוגות המכילים את התיאור החיובי וכל הצילומים(!!) כאשר השני מכיל את המרחקים בין הצילום החיובי לכל התיאורים מהמיני-באטץ. הלוס הסופי הינו סכום של שני הלוסים האלו.

הישגי מאמר: בוצע אימון pretrain של ConViRT על שני דיאטה סטים: MIMIC-CXR, הדאטה סט musculoskeletal מ rhode island hospital. אחר כך הם כיילו את המודל שלהם לכמה סוגים של משימות:

1. סיווג צילום,

דאטה סטים: RSNA Pneumonia Detection, CheXpert, CovidX, MURA

2. מציאת צילום הדומה ביותר לצילום נתון (Zero-shot Image-image Retrieval)

דאטה סטים: CheXpert 8×200 Retrieval Dataset.

3. מציאת צילום הכי דומה לתיאור נתון (דאטה סט כמו הקודם)

בכל המשימות האלו הם הצליחו להשיג ביצועים טובים ממגוון שיטות pretraining

לינק למאמר: **https://arxiv.org/abs**/2010.00747

לינק לקוד: לא הוגש לארקיב

נ.ב. מאמר עם רעיון מגניב להתגבר על קושי בבניית ייצוגי התמונות בדומיין הרפואי. כתוב מאוד ברור ומפורט. מומלץ