R-CNN(Regions CNN)

מאמר: https://arxiv.org/pdf/1311.2524.pdf

R-CNN משמש לסיווג ולאיתור אובייקטים עם bbox עבור מספר עצמים שנמצאים בתמונה.
ב R-CNN במקום להריץ על התמונה מספר עצום של סיווגים (sliding window), אנחנו מעבירים את התמונה דרך selective search ולוקחים את ה 2000 region proposal מהתוצאה ומריצים עליהם את המסווג. בדרך זו במקום לסווג מספר עצום של אזורים אנו צריכים רק לסווג את 2000 האזורים הראשונים. זה הופך את האלגוריתם הזה למהיר בהשוואה לטכניקות קודמות לגילוי אובייקטים.

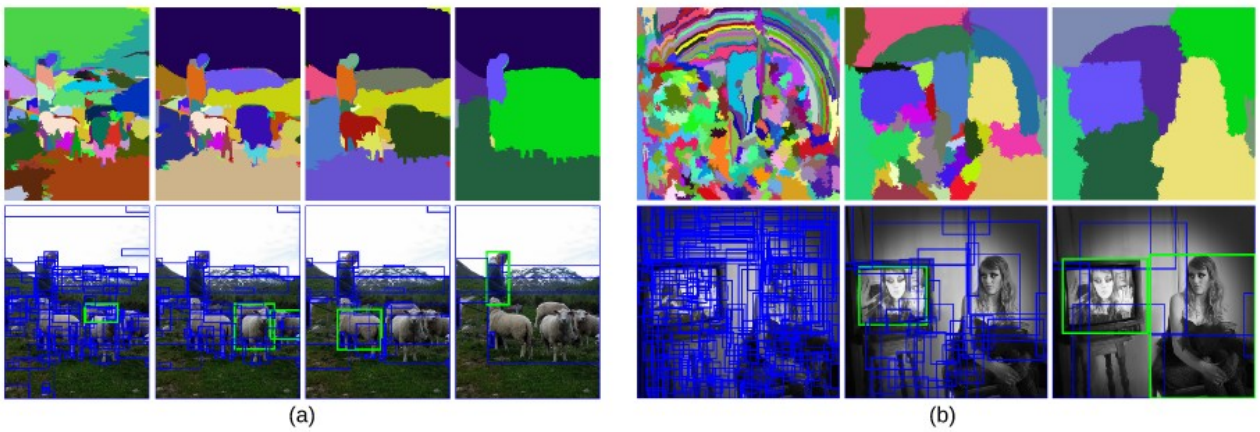
שלבי העבודה של R-CNN

1. צור category-independent region proposals על ידי שימוש באלגוריתם selective search כדי לחלץ 2000 region proposals.
2. Region proposals מוזנים לרשת CNN. רשת ה CNN משמשת כמחלץ תכונות ומוציא כפלט וקטור תכונות באורך קבוע. לאחר שעבר ב CNN, R-CNN מחלץ וקטור פיצ'רים באורך 4096 עבור כל reggion proposal.
3. החל SVM עבור על וקטור תכונות שחולצו מרשת ה CNN. באמצאות SVM אנו מבצעים את הסיווג של האזורים המוצעים. Regressor משמש לחיזוי הקואורדינטות של ה bbox.

עבור כל האזורים המוצעים בתמונה, נבצע non-maximum suppression גרידי על מנת להמנע מאזורים חופפים.

Region proposal

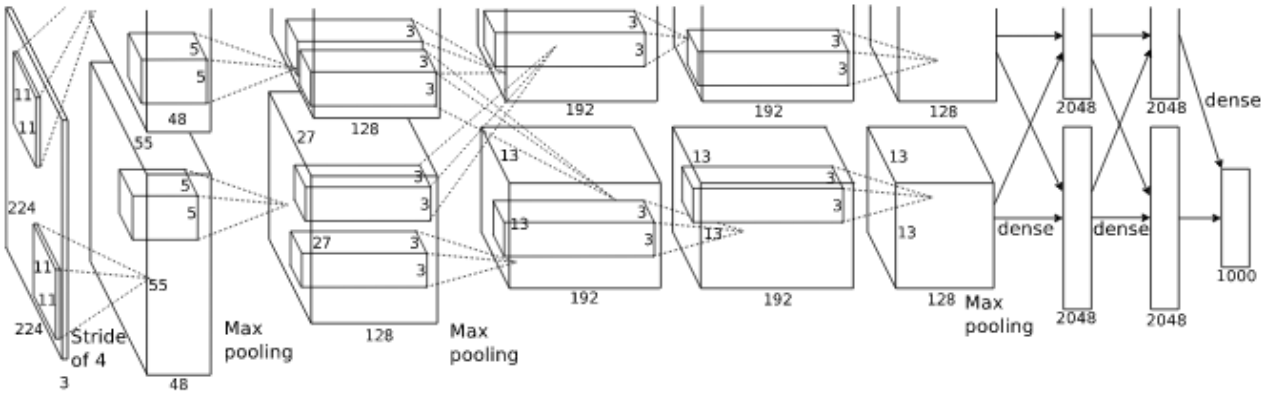
השלב הראשון ב RCNN הוא למצוא אזורים בתמונה שעשויים להיות שייכים לאובייקט מסויים. כותבי המאמר השתמשו ב selective search אלגורים על מנת למצוא אזורים אלו. במאמר משתמשים ב selective search כדי לייצר 2000 region proposals עבור כל תמונה.



http://www.huppelen.nl/publications/selectiveSearchDraft.pdf

Feature extraction

בסוף שלב זה, עבור כל region proposal, נוצר וקטור בגודל 4096 המכיל את הפיצ'רים שלו על ידי CNN. עבור ה CNN המחברים השתמשו ב AlexNet, המכיל 5 שכבות קונבולוציה ו 2 שכבות FC.



Supervised Pre-training: ה CNN אומן על ILSVRC2012 classification dataset.

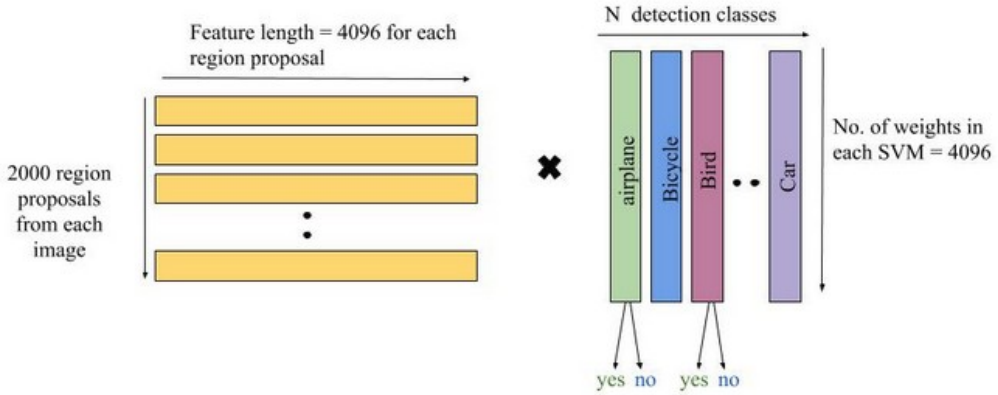
Domain-Specific Fine-Tuning: כעת נבמע transfer learning על מנת ללמד את הרשת לחלץ את התכונות מה region proposals. בנוסף, מכוננים את רשת הסיווג כדי לזהות את הכיתות השייכות למשימת הגילוי.

SVM for object classification

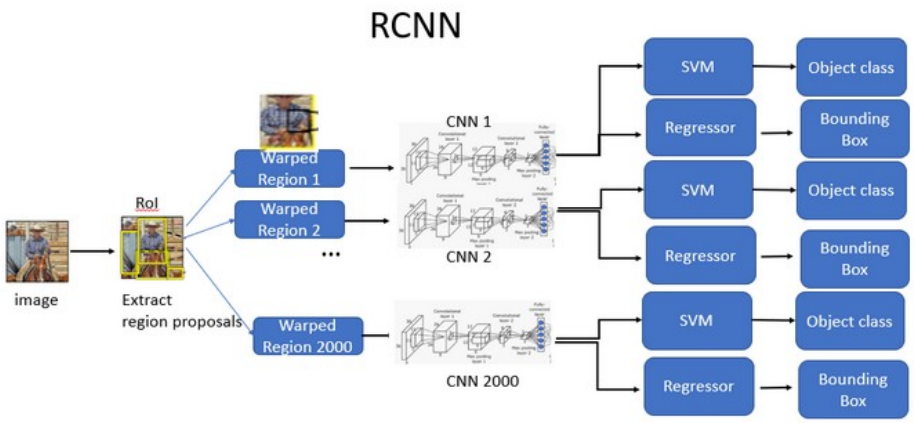
שלב זה מורכב מלמידה של מסווג SVM לינארי יחיד.

קלט: וקטור באורך 4096 לכל אזור מוצע.

אופן פעולה: עבור כל תמונה נוצר מטריצת מאפיינים של 2000X 4096 . מטריצת המשקל SVM היא 4096X N כאשר N הוא מספר המחלקות.



Architecture of R-CNN



אופן פעולה:

- חילוץ 2000 אזורים על ידי שימוש באלגוריתם selective search.
- חילוץ תכונות על ידי CNN עבור כל אזור בתמונה, עבור כל תמונה. כלומר עבור N תמונות, יהיה לנו N*2000 CNN וקטורי פיצ'רים.
- R-CNN מאתרת אובייקטים בשלושה שלבים:
  - CNN לחילוץ תכונות.
  - SVM לינארי לסיווג האובייקט.
  - Regression לשם חיזוי ה bbox.

הפלט של R-CNN:עבור כל אזור מוצע שהתגלה בו אובייקט על ידי ה SVM, יש לנו bbox מסביב לאובייקט.