# Dual Encoding for Zero-Example Video Retrieval

## Problema

Recuperación de video sin ningún ejemplo es un problema donde se tiene una query descrita en texto de lenguaje natural y se intenta de obtener frames del video que tengan referencia/relevancia a la query, esto sin tener videos con los tags que lo represente.

También se da la inversa, tener frames de videos y obtener una oración que describa dichos frames, por consiguiente el “*dual encoding*”.

## Propuesta de solución del paper

La propuesta tiene el objetivo de generar un método libre de contexto que entienda la semántica para ambos, videos y oraciones.

Proponen un encoding dual de múltiples niveles, donde la combinación de los niveles permite el manejo de ambas modalidades (texto y video).

El primer nivel tiene el objetivo de extraer *features* del video, utilizando Mean Pooling que calcula un promedio de los features del video.

El segundo nivel utiliza un Bidirectional GRU (biGRU) para modelar la información temporal del video.

El tercer nivel construye CNN sobre el biGRU para mejorar los encodings locales. Esto permite la diferenciación entre vídeos con diferencias sutiles.

El cuarto nivel utiliza VSE++ para llevar a un espacio común las representaciones de video y texto.

Entrega Parcial 1

Descripción detallada del problema que enfrenta el artículo: deben indicar la motivación del artículo, el problema que trata de resolver, por qué el problema es relevante y definirlo como un problema de Aprendizaje de Máquina (definido como input y output).

Descripción detallada de las métricas: Deben describir las métricas que usan en el artículo para medir el desempeño de la solución. No solo deben enumerar los nombres de las métricas si no investigar a qué se refiere cada una y definirla formalmente (muchos artículos usan métricas que van mucho más allá de acierto o precisión). Deben además decidir qué métricas son las que usarán ustedes en su replicación. Algunos artículos usan muchas métricas distintas. En su caso, basta con que seleccionen un par de las mencionadas en el artículo.

Descripción de los datos usados: Deben investigar qué datos usaron los autores para medir la eficacia de su solución, deben descargarlos para poder analizarlos, describirlos (tamaño, tipo de datos, composición, etc.) y, siempre que sea posible, dejarlos en su repositorio, o de otra forma indicar cómo se deben descargar y procesar. Como plus, deben construir un DataSet y un DataLoader para los datos (o buscar si es que existe alguno disponible en la Web).

Descripción de la arquitectura: Deben describir la arquitectura de Deep Learning usada por los autores para resolver el problema. Idealmente entender y enumerar las fórmulas usadas en la arquitectura. Si las fórmulas o la arquitectura son muy complicadas, pueden sólo incluir una descripción de alto nivel, pero procuren juntarse con su tutor para que los guíe en el entendimiento de la arquitectura.

Repositorio de código: Finalmente deben indicar un repositorio (link Web, idealmente github) en donde irán dejando su código. Para esta entrega no es necesario que tengan código funcional pero sí que el repositorio esté creado. Incentivamos comenzar construyendo código tan pronto como sea posible. Si tienen dudas de qué código deben construir y cuál pueden usar desde repositorios públicos, consúltenlo con su tutor.