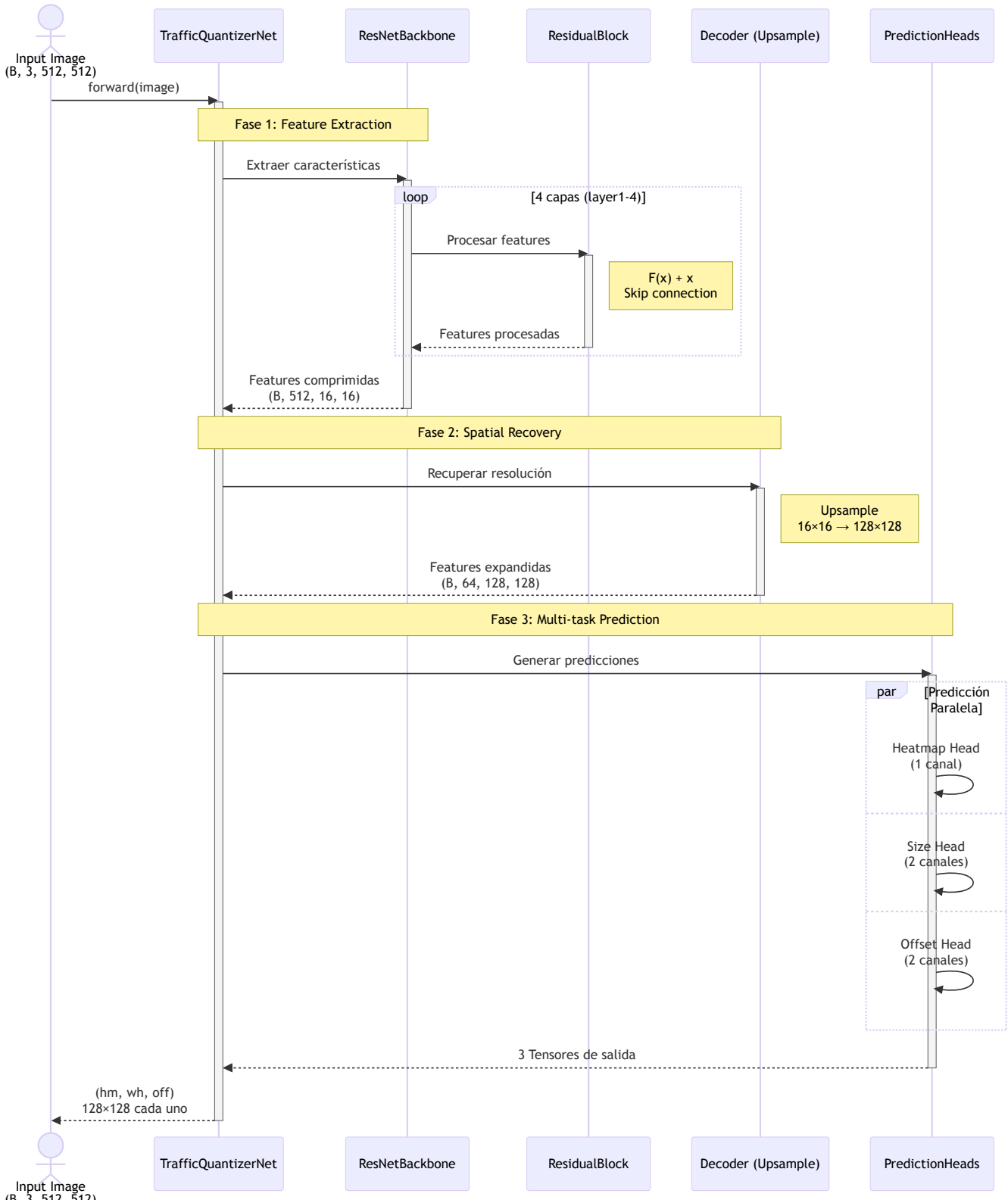
Diagrama de Secuencia -

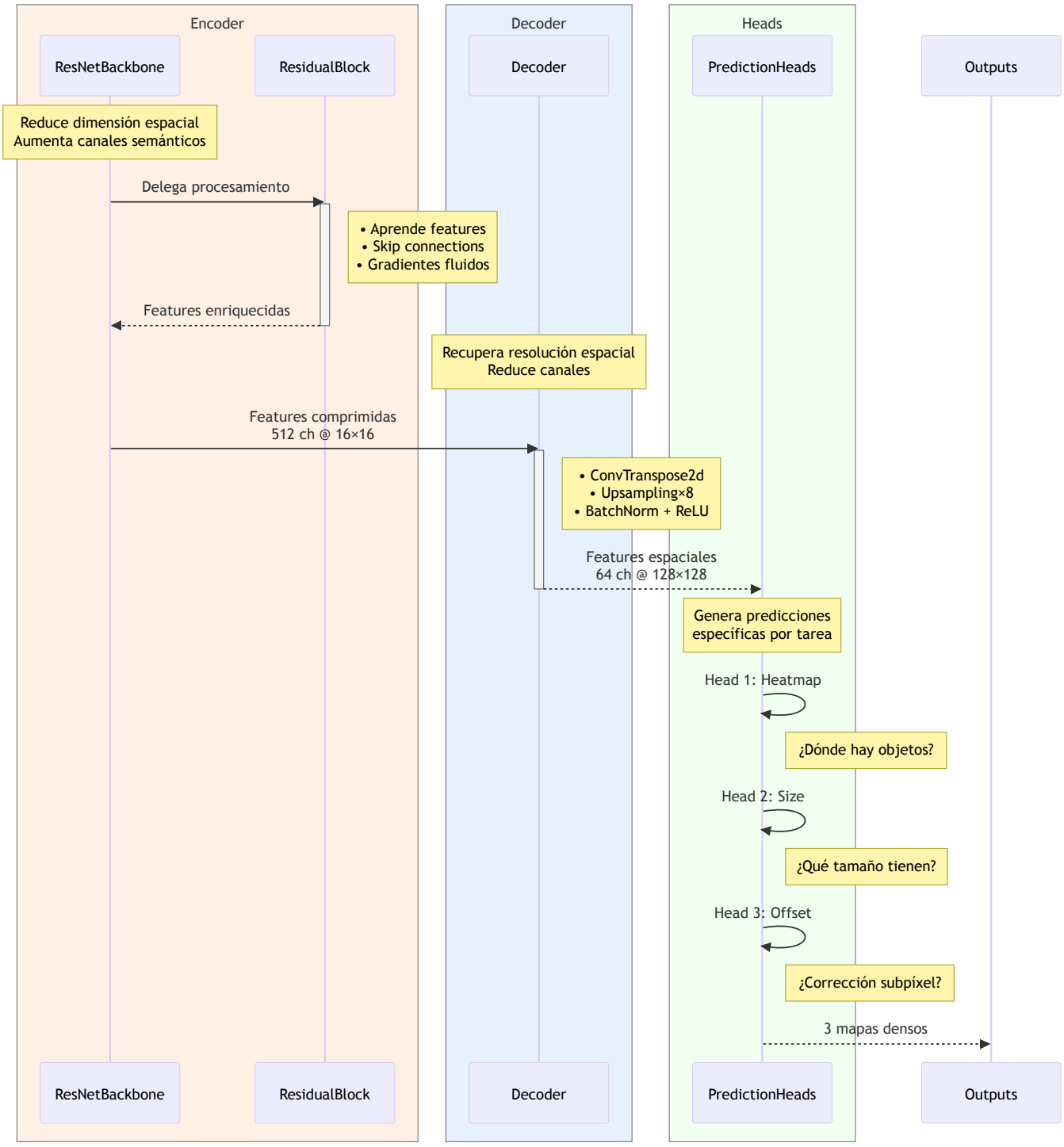
TrafficQuantizerNet

Diagrama Simplificado (Alto Nivel)

Interacción entre Componentes Principales


Input Image
(B, 3, 512, 512)
TrafficQuantizerNet
ResNetBackbone
ResidualBlock
Decoder (Upsample)
PredictionHeads
forward(image)
Fase 1: Feature Extraction
Extraer características
loop
[4 capas (layer1-4)]
Procesar features
F(x) + x
Skip connection
Features procesadas
Features comprimidas
(B, 512, 16, 16)
Fase 2: Spatial Recovery
Recuperar resolución
Upsample
16×16 → 128×128
Features expandidas
(B, 64, 128, 128)
Fase 3: Multi-task Prediction
Generar predicciones
par
[Predicción Paralela]
Heatmap Head
(1 canal)
Size Head
(2 canales)
Offset Head
(2 canales)
3 Tensores de salida
(hm, wh, off)
128×128 cada uno
Input Image
TrafficQuantizerNet
ResNetBackbone
ResidualBlock
Decoder (Upsample)
PredictionHeads

Rol de Cada Componente



Función de Cada Clase

Clase	Responsabilidad	Input	Output
ResNetBackbone	Extracción de características Compresión espacial	(B, 3, 512, 512)	(B, 512, 16, 16)

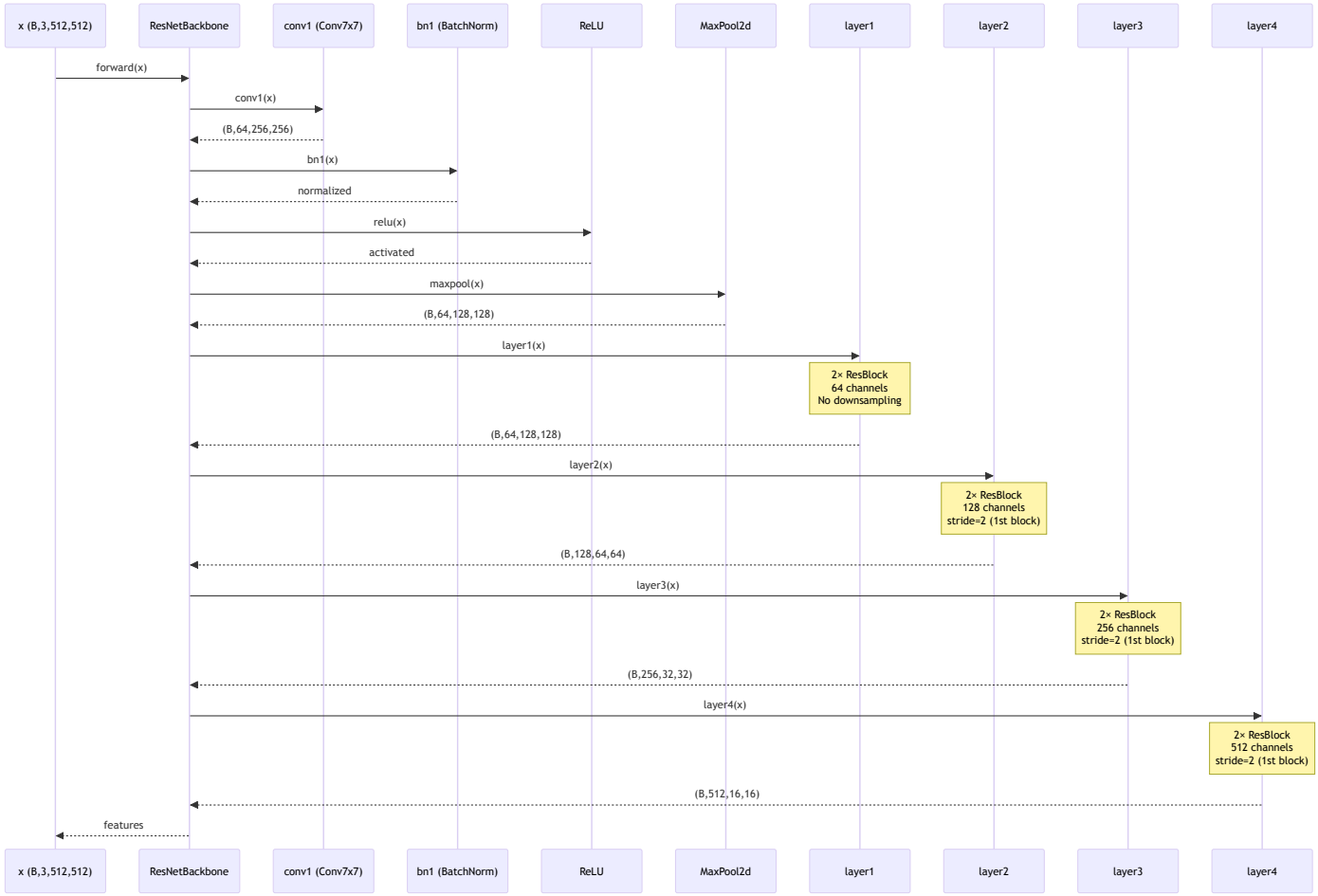
Clase	Responsabilidad	Input	Output
ResidualBlock	Transformación no lineal con skip connection	(B, C_in, H, W)	(B, C_out, H', W')
Decoder	Recuperación espacial Upsampling progresivo	(B, 512, 16, 16)	(B, 64, 128, 128)
PredictionHeads	Predicciones multi-tarea paralelas	(B, 64, 128, 128)	3× (B, C, 128, 128)

Lógica Interna del Modelo (Detallada)

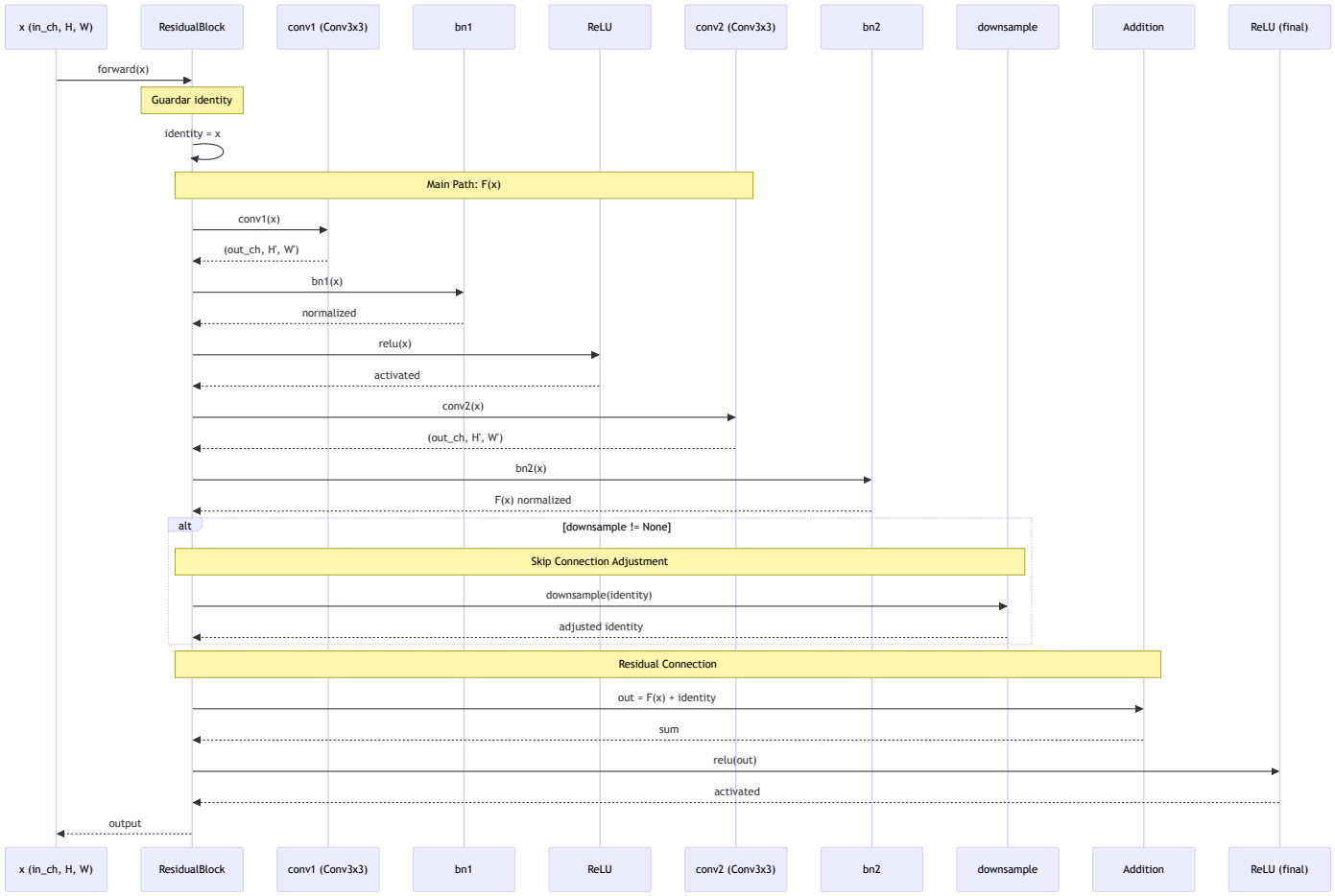
Secuencia de Forward Pass Completo

Flujo Detallado por Componente

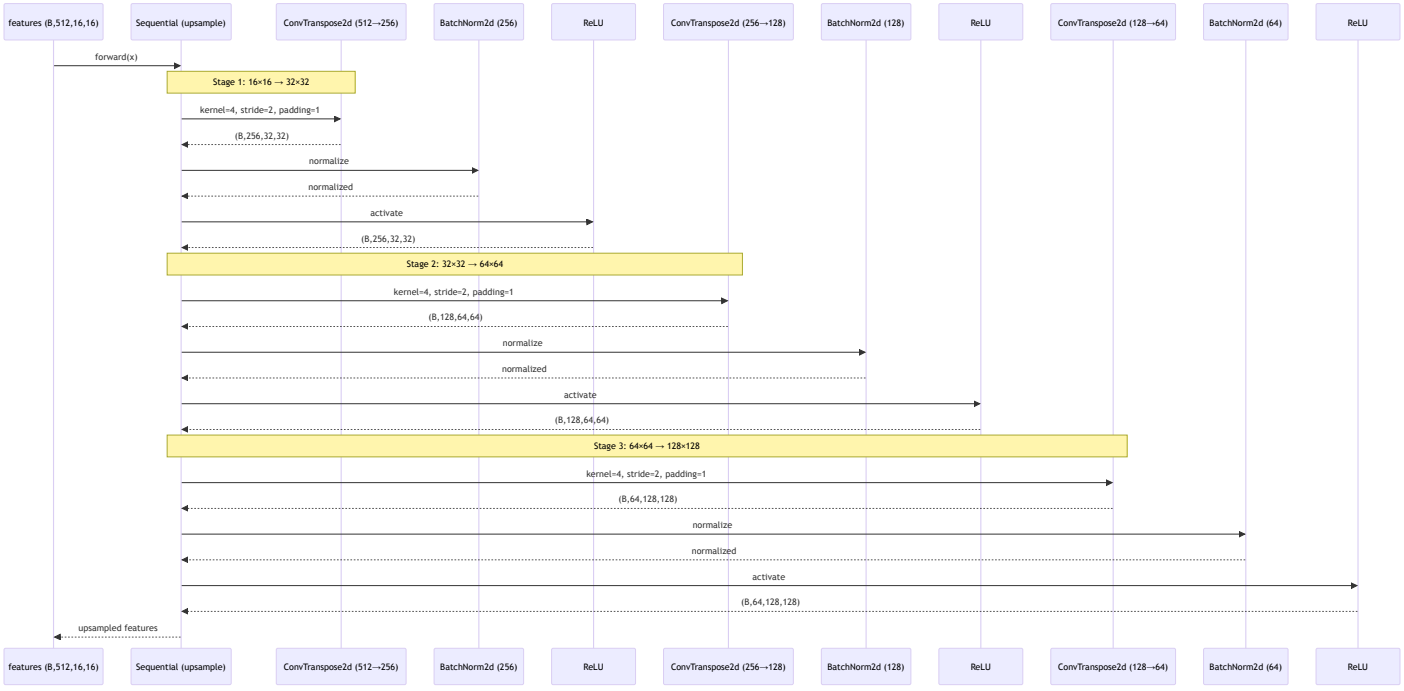
1. Backbone - ResNetBackbone



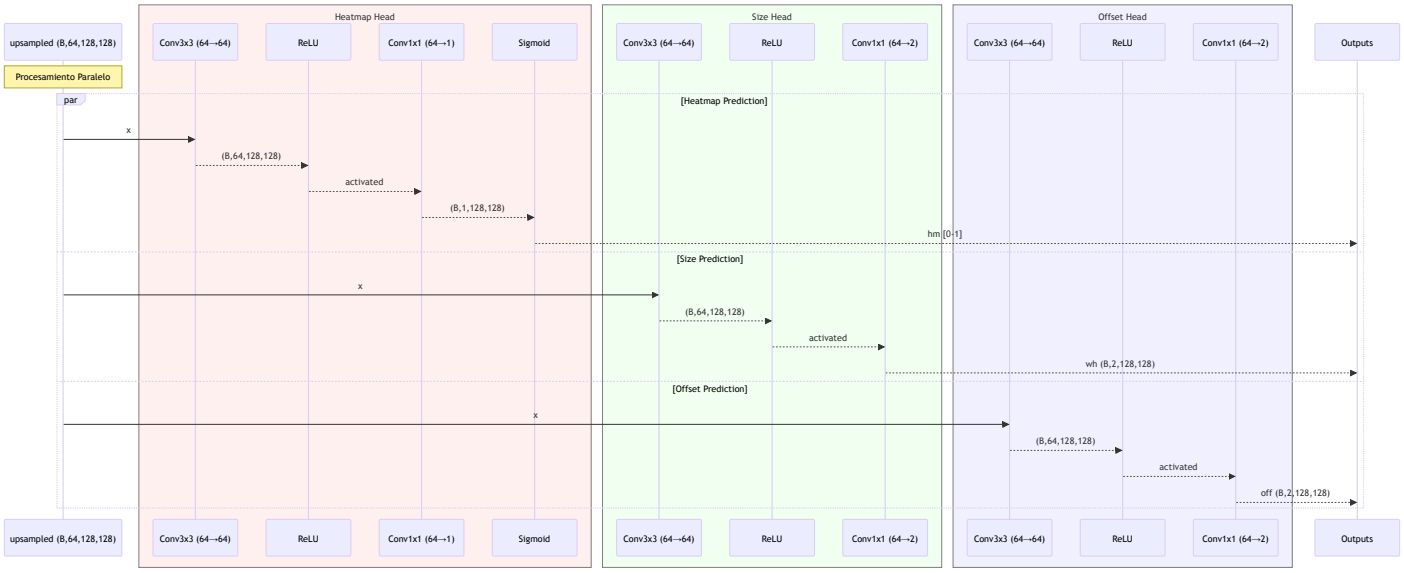
2. ResidualBlock - Bloque Individual



3. Decoder - Upsample Module



4. Prediction Heads (Paralelos)



Transformaciones de Dimensiones

Tabla de Cambios Espaciales

Etapa	Operación	Input Shape	Output Shape	Cambio
Input	-	(B, 3, 512, 512)	-	-
Stem Conv	Conv7×7 s=2	(B, 3, 512, 512)	(B, 64, 256, 256)	÷2
Stem Pool	MaxPool s=2	(B, 64, 256, 256)	(B, 64, 128, 128)	÷2
Layer 1	2× ResBlock	(B, 64, 128, 128)	(B, 64, 128, 128)	=
Layer 2	2× ResBlock	(B, 64, 128, 128)	(B, 128, 64, 64)	÷2
Layer 3	2× ResBlock	(B, 128, 64, 64)	(B, 256, 32, 32)	÷2
Layer 4	2× ResBlock	(B, 256, 32, 32)	(B, 512, 16, 16)	÷2
Decoder 1	ConvT 512→256	(B, 512, 16, 16)	(B, 256, 32, 32)	×2
Decoder 2	ConvT 256→128	(B, 256, 32, 32)	(B, 128, 64, 64)	×2
Decoder 3	ConvT 128→64	(B, 128, 64, 64)	(B, 64, 128, 128)	×2
Head HM	Conv→Sigmoid	(B, 64, 128, 128)	(B, 1, 128, 128)	=

Etapa	Operación	Input Shape	Output Shape	Cambio
Head WH	Conv→Linear	(B, 64, 128, 128)	(B, 2, 128, 128)	=
Head OFF	Conv→Linear	(B, 64, 128, 128)	(B, 2, 128, 128)	=

Resumen del Flujo

```
Input: 512×512 (RGB)
    ↓ Stem (÷4)
128×128 (64 ch)
    ↓ Layer1 (=)
128×128 (64 ch)
    ↓ Layer2 (÷2)
64×64 (128 ch)
    ↓ Layer3 (÷2)
32×32 (256 ch)
    ↓ Layer4 (÷2)
16×16 (512 ch) ← Bottleneck
    ↓ Decoder (×8)
128×128 (64 ch)
    ↓ 3 Heads
128×128 (5 ch total)
    • Heatmap: 1 ch
    • Size: 2 ch
    • Offset: 2 ch
```

Tiempo de Ejecución Estimado

Componente	Operaciones	Tiempo Relativo
Stem	Conv7×7 + Pool	~5%
Layer1-4 (Backbone)	8 ResBlocks	~60%
Decoder	3 ConvTranspose2d	~25%
Heads	3× (Conv3×3 + Conv1×1)	~10%

Total Forward Pass: ~15-30ms en Apple M1/M2 (batch=1)